Production Function Estimation with Factor-Augmenting Technology: An Application to Markups

Mert Demirer

May 13, 2025

Abstract

I develop a new method for estimating production functions with labor-augmenting technology and apply it to markup estimation. The method imposes limited parametric restrictions and generalizes prior approaches that rely on the CES production function. I first extend the canonical Olley-Pakes framework assumptions to multidimensional productivity and then develop an identification strategy based on a control variable approach and first-order conditions. I use this method to estimate output elasticities and markups in manufacturing industries in the US and four developing countries. I find that neglecting labor-augmenting productivity overestimates capital elasticity, underestimates variable input elasticity, and overestimates markups in all countries.

Keywords: Production Functions, Markups, Factor-Augmenting Technology, Control Variables, Productivity, Manufacturing

---

Assistant Professor, MIT Sloan School of Management, mdemirer@mit.edu. I am indebted to Nikhil Agarwal, Daron Acemoglu, Anna Mikusheva, and Victor Chernozhukov for invaluable guidance and support. This paper benefited from feedback from Daniel Ackerberg, Allan Collard-Wexler, Glenn Ellison, Sara Fisher Ellison, Paul Joskow, Christopher Knittel, Jing Li, Whitney Newey, Alex Olssen, Nancy Rose, Tobias Salz, Karthik Sastry, John Van Reenen, Michael Whinston, Kamil Yilmaz, and Nathan Zorzi, as well as various seminar participants. I thank the State Institute of Statistics in Turkey for allowing me access to the data under the project "Growth and Technological Change in Turkey." I thank Devesh Raval for helping me to obtain the Colombian and Chilean data.

1 Introduction

Production functions play a central role in many areas of economics, including research on firm productivity, input misallocation, and market power (Syverson 2011; Hsieh and Klenow 2009; De Loecker et al. 2020). Studying these topics typically involves specifying a production function model and estimating it using firm-level data. However, misspecified production functions can produce biased estimates, which in turn lead to incorrect conclusions about various economic questions. For example, biased capital elasticities might falsely suggest misallocation in an economy with efficient allocation (Haltiwanger et al. 2018), and biased flexible input elasticities might generate incorrect markup estimates (Raval 2023).

Much of the empirical literature relies on factor-neutral productivity and parametric functional forms to estimate production functions (De Loecker and Syverson 2021). These assumptions impose theoretical restrictions that may not adequately capture heterogeneity in firms' production technologies. For instance, under factor-neutral technology, productivity shocks do not generate any unobserved heterogeneity in output elasticities. Similarly, the widely used Cobb-Douglas specification imposes a unitary elasticity of substitution among all input pairs.

In this paper, I develop a method to estimate gross production functions with factor-augmenting productivity, and I examine its implications empirically. The production function specification has two main features. First, it incorporates labor-augmenting productivity in addition to the standard Hicks-neutral productivity. Second, it requires a minimal functional form assumption known as homothetic separability. Together, these features provide a flexible framework that aims to capture heterogeneity in production technologies across firms.

To study factor-augmenting production functions, I make both methodological and empirical contributions. Methodologically, I extend the canonical production function estimation assumptions from Olley and Pakes (1996) to accommodate a model with multidimensional productivity, and analyze this model's identification building on the existing literature (Doraszelski and Jaumandreu 2018; Raval 2019). Empirically, I find that neglecting labor-augmenting productivity introduces biases in output elasticities and markups across several datasets.

The paper establishes three main results in developing its identification strategy. First, I show that, under cost minimization, labor-augmenting productivity can be expressed as a function of inputs by inverting firms' flexible input demand functions. This result extends the approach of Doraszelski and Jaumandreu (2018) to a nonparametric framework and enables controlling for labor-augmenting productivity. The key assumption underlying this result is homothetic separability, a condition that accommodates many of the commonly used parametric production functions (Shephard 1953; Clemhout 1968). Importantly, it is an economic rather than a statistical restriction with clear implications for firm behavior.

Second, I develop a control variable approach for two-dimensional productivity, building on Imbens and Newey (2009). This approach exploits the timing assumption for capital input and the Markov property of productivity shocks, both of which are standard assumptions in the production function literature (Ackerberg et al. 2015; Ackerberg et al. 2023). To implement the Imbens and Newey (2009) approach, I show that under the modeling assumptions, the input demand functions exhibit a triangular structure with respect to productivity shocks. This structure overcomes the invertibility problems typically encountered in models with multidimensional unobserved heterogeneity (Kasy 2011).

The third result provides an identification strategy for output elasticities. After developing my control variable approach, I investigate which aspects of the production function can be identified using input and output data. I first establish that only the sum of the flexible input elasticities—rather than labor and materials elasticities separately—can be identified from variation in inputs and output due to a functional dependence problem. To separately identify these elasticities, I leverage the first-order conditions (FOCs) of firms' cost minimization problem, which imply that the ratio of the elasticities of two flexible inputs equals the ratio of their revenue shares. Therefore, incorporating data on revenue shares of labor and materials enables separate identification of the individual elasticities.

A notable advantage of using FOCs to identify the ratio of output elasticities is that the resulting markup estimates from two different flexible inputs are identical. This feature addresses recent evidence that different flexible inputs often yield conflicting markup estimates in Hicks-neutral production functions (Doraszelski and Jaumandreu 2019; Raval 2023). I show that incorporating labor-augmenting productivity into the production function provides a solution to this discrepancy.

Models with only Hicks-neutral productivity may lack the flexibility needed to capture variation in input ratios across firms, leading to conflicting markup estimates from different inputs. Labor-augmenting productivity introduces an additional dimension of unobserved heterogeneity and makes the model internally consistent.

In my empirical applications, I apply the proposed method to estimate output elasticities in manufacturing industries using Compustat data for the US and plant-level data for Chile, Colombia, India, and Turkey. I compare the estimates against three alternative production functions with varying levels of flexibility in terms of functional form and unobserved heterogeneity: (i) Cobb-Douglas, (ii) Translog with only Hicks-neutral productivity, and (iii) CES with both Hicks-neutral and labor-augmenting productivity. The results suggest that the Cobb-Douglas model underestimates capital elasticity by 25.8% and overestimates flexible input elasticity by 11.1% on average. Although CES and translog specifications reduce these biases, neither fully eliminates them, highlighting the importance of both relaxing parametric restrictions and incorporating unobserved heterogeneity.

Output elasticities often serve as key inputs for measuring various economic outcomes. A prominent example is the production approach to markup estimation, which has gained popularity in recent years (De Loecker and Warzynski 2012; De Loecker et al. 2020). In this approach, markup is given by the output elasticity of a flexible input divided by that input's revenue share. Having documented biases in elasticity estimates, I analyze how these biases propagate into markup estimates.

I find that the Cobb-Douglas model systematically overestimates aggregate markups by 6.3 to 13.4 percentage points (pp) across countries compared to the homothetic labor-augmenting production function. This difference persists, though smaller in magnitude, with more flexible specifications than Cobb-Douglas, such as the Hicks-neutral translog or labor-augmenting CES production functions. A decomposition analysis suggests two key drivers behind these results: (i) the Cobb-Douglas model overestimates the average level of flexible input elasticity in the economy, and (ii) it fails to capture the inverse relationship between firm size and flexible input elasticity, thereby leading to higher markup biases for larger firms.

Next, I examine how estimating a labor-augmenting production function affects changes in markups by analyzing markup trends in US manufacturing. The results suggest that the aggregate markup increased by 11.7 pp—from 1.26 in the 1960s to

1.38 in the 2010s—as opposed to the 25.2 pp increase implied by the Cobb-Douglas specification. This result arises because the Cobb-Douglas model suggests minimal changes in production technology over time, whereas the labor-augmenting production function indicates larger changes in flexible input elasticities.

Although this paper introduces unobserved heterogeneity in production functions, several limitations remain. First, the model does not allow for unobserved demand shocks, as they increase the dimensionality of unobserved heterogeneity and violate the invertibility of input demand (Ackerberg and De Loecker 2024). Despite this limitation, I show that the model is consistent with a broad range of demand systems that follow an aggregative game structure (Nocke and Schutz 2018). Second, the paper does not address other key challenges in production function estimation, such as multi-output firms, unobservability of input prices, and distortions in input markets (De Loecker and Syverson 2021). Third, as Bond et al. (2021) note, markup estimation requires data on physical output rather than revenues, which are available only in one of my datasets. Nevertheless, by leveraging the limited data on quantities and input prices as well as simulation exercises, I provide empirical evidence suggesting that these concerns are unlikely to drive the main patterns I observe when comparing production function models.

Contribution to the Literature. This paper extends the literature on production function estimation with proxy variables by integrating factor-augmenting productivity into the production function (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg et al. 2015; Gandhi et al. 2020). My approach builds on these studies, but differs from them by using the FOCs of cost minimization and employing the control variable method of Imbens and Newey (2009).[1] The main distinction in my use of FOCs compared to Gandhi et al. (2020) is that I require two flexible inputs rather than one. Additionally, this paper relates closely to Ackerberg et al. (2023) and Pan (2024), who apply the control variable approach to production functions with a single but fully non-separable productivity shock.

Three recent papers have studied factor-augmenting technology and its implications (Doraszelski and Jaumandreu 2018; Raval 2019; Zhang 2019). Common

[1]The proxy and control variable approaches are sometimes used interchangeably in the literature. In this paper, the proxy variable approach refers to the method of Ackerberg et al. (2015), while the control variable approach refers to the Imbens and Newey (2009) approach.

features of these papers are the CES production function and firm-level variation in input prices. They exploit parameter restrictions between the production and input demand functions, inverting the latter to recover factor-augmenting productivity. I contribute to this literature by relaxing the CES assumption and analyzing identification with or without input price variation.[2]

This paper also relates to the growing literature that studies the properties of the production approach to markup estimation. Raval (2023) finds that different flexible inputs produce conflicting markup patterns and suggests labor-augmenting productivity as a potential solution. Using US Census data, Foster et al. (2024) find evidence that production function estimation at a more granular level leads to smaller increases in markups over time. Doraszelski and Jaumandreu (2023) emphasize the importance of controlling for demand in production function estimation. Finally, Bond et al. (2021) show that revenue data alone does not identify markups, while De Ridder et al. (2025) argue that such data can inform markup dispersion under certain assumptions. I contribute to this literature by analyzing how allowing for labor-augmenting technology affects markup estimation.

2 Production Function Model

This section introduces a production function model and provides its assumptions.

2.1 Production Function with Labor-Augmenting Technology

Firm i produces output at time t using three inputs—capital, K_{it}; labor, L_{it}; and materials, M_{it}—according to the following production function:

$$Y_{it} = F_t(K_{it}, \omega_{it}^L L_{it}, M_{it}) \exp(\omega_{it}^H) \exp(\epsilon_{it}), \tag{2.1}$$

where Y_{it} denotes the output produced by the firm. The production function is *industry-specific* and *time-varying*, meaning that firms in the same industry share a common functional form that can change over time. The production function includes two unobserved productivity terms: labor-augmenting productivity, $\omega_{it}^L \in \mathbb{R}_+$, which increases the effective unit of labor input, and Hicks-neutral productivity, $\omega_{it}^H \in \mathbb{R}$, which raises the output level for any input composition. Finally, $\epsilon_{it} \in \mathbb{R}$ is a random shock to output.

[2] Another strand of literature uses a random coefficient model to introduce firm-level unobserved heterogeneity in output elasticities (Li and Sasaki 2017; Balat et al. 2022; Kasahara et al. 2023).

I assume that labor and materials are *flexible inputs*, meaning that the firm chooses them each period after observing its productivity shocks, and their levels do not affect future production.[3] Capital, in contrast, is a *predetermined input*, chosen one period ahead of production. Each period, the firm chooses its flexible inputs to minimize production costs for a profit-maximizing target output, given its information set $\mathcal{I}_{it}$. The information set includes productivity, capital stock, and other relevant variables observed by the firm. The shock ϵ_{it} is independent of the information set and can be interpreted as an ex-post productivity shock.

I assume that firms are price-takers in the input market.[4] In the output market, the model allows for imperfect competition but imposes certain restrictions on product market competition to reduce the dimensionality of unobserved heterogeneity, as discussed later in this section. The baseline model assumes uniform input prices across firms for labor and materials (p^l_t, p^m_t), while Online Appendix A extends the model to include heterogeneous input prices.

One limitation of the model is that it incorporates factor-augmenting productivity only for labor, implying that there are no capital- or materials-specific productivity terms in the production function. More broadly, the results in this paper accommodate factor-augmenting productivity for only one flexible input. I focus on labor-augmenting productivity because its variation can capture cross-firm differences in labor productivity due to various factors, such as differences in firms' management practices (Bloom and Van Reenen 2010) and human capital (Dunne et al. 1997). Since these factors are often unobserved in production datasets, it is natural to model them as unobserved heterogeneity.

Labor-augmenting productivity relaxes certain restrictions on firms' production technology imposed by Hicks-neutral production functions. For example, in the case of the Cobb-Douglas form, revenue shares of flexible inputs must be uniform across firms (Jorgenson 1986). Even more flexible functional forms (e.g., the

[3]The flexible labor input assumption may be strong due to potential adjustment costs. Whether labor is a flexible input depends on its measurement (e.g., hours worked vs. employees), specific industry, and country, so the plausibility of this assumption depends on the specific setting. In my empirical application to manufacturing, where the workforce mainly consists of production workers, this assumption is more plausible than in industries dominated by white-collar workers.

[4]As observed by Rubens et al. (2025), this assumption is critical because otherwise, firm-level markdowns introduce additional unobserved heterogeneity, making identification of factor-augmenting productivity challenging.

Hicks-neutral translog) can be restrictive as they do not allow for unobserved variation in output elasticities. Accounting for this variation is important because the literature has shown large heterogeneity across firms and a decline over time in labor share in many economies. These trends are linked to within-industry changes (Autor et al. 2020; Kehrig and Vincent 2021), and heterogeneity in production technology is proposed as a mechanism (Oberfield and Raval 2021).

2.2 Assumptions

This section presents assumptions and discusses their implications. The first assumption introduces a homothetic separability restriction, while the subsequent assumptions pertain to firm behavior and productivity shocks, extending the standard Markov and monotonicity assumptions to a setting with two-dimensional productivity. Throughout the paper, I assume that all functions are continuously differentiable as needed and all random variables have continuous and strictly increasing distribution functions.

2.2.1 A Homothetic Separability Restriction

I start by providing a set of conditions under which labor-augmenting productivity can be expressed as a function of firms' inputs. Let $\sigma_t(K_{it}, \omega_{it}^L L_{it}, M_{it})$ represent the elasticity of substitution between labor and materials.[5]

Assumption 2.1 (Homothetic Separability)**.** *Suppose that:*

(i) *The production function satisfies the following functional form*

$$Y_{it} = F_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big) \exp(\omega_{it}^H) \exp(\epsilon_{it}). \tag{2.2}$$

(ii) $h_t(K_{it}, \cdot, \cdot)$ *is homogeneous of arbitrary degree for all* K_{it}*.*

(iii) $\sigma_t(K_{it}, \omega_{it}^L L_{it}, M_{it})$ *is everywhere greater than 1 or everywhere less than 1.*

Assumption (i-ii) is called homothetic separability (Shephard 1953), and it is the key assumption of the paper. It states that the production function is separable in K_{it} and a composite input $h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})$ that is homogeneous of arbitrary degree in labor and materials. Homothetic separability commonly appears in models of consumer preferences and production functions (Lewbel and Linton 2007). It has two main economic implications. First, production can be

[5]The formal definition of elasticity of substitution appears in Equation (A.7) of Appendix A.

conceptualized in two stages, where $h_t(\cdot)$ is treated as an 'intermediate input' with its own production function. Second, the homotheticity of $h_t(\cdot)$ implies that jointly scaling labor and materials inputs is equivalent to scaling $h_t(\cdot)$. Therefore, instead of jointly optimizing labor and materials, the firm can first choose the optimal materials-to-labor ratio and then determine the optimal scale of $h_t(\cdot)$. This two-stage optimization simplifies the firm's cost minimization problem.

Assumption 2.1(iii) states that labor and materials are either everywhere substitutes or everywhere complements. In a nonparametric production function, the degree of substitutability among inputs can vary with their levels, so this assumption rules out cases where labor and materials switch between substitutes and complements. Next, I provide two parametric forms that satisfy Assumption 2.1.

Examples (CES and Nested CES). The CES production function is given by:

$$Y_{it} = \left(\beta_k K_{it}^{\rho} + \beta_l[\omega_{it}^L L_{it}]^{\rho} + (1-\beta_k-\beta_l)M_{it}^{\rho}\right)^{v/\rho} \exp(\omega_{it}^H)\exp(\epsilon_{it}).$$

Assumption 2.1 nests the CES production function with $h_t(\cdot) = \beta_l\left[\omega_{it}^L L_{it}\right]^{\rho} + (1-\beta_k-\beta_l)M_{it}^{\rho}$, which is homogeneous of degree one and has elasticity of substitution $\sigma = 1/(1-\rho)$. The CES specification has been widely used in the literature to study factor-augmenting technology (Doraszelski and Jaumandreu 2018). A more flexible functional form with homothetic separability is the nested CES:

$$Y_{it} = \left(\beta_k K_{it}^{\rho} + (1-\beta_k)\left(\beta_l\left[\omega_{it}^L L_{it}\right]^{\rho_1} + (1-\beta_l)M_{it}^{\rho_1}\right)^{\rho/\rho_1}\right)^{v/\rho} \exp(\omega_{it}^H)\exp(\epsilon_{it}),$$

where labor and materials are nested with elasticity of substitution $1/(1-\rho_1)$.[6,7]

Assumption 2.2 (Cost Minimization)**.** *The firm minimizes static production costs of producing target output $\bar{Y}_{it}$ with respect to (L_{it}, M_{it}) given K_{it}, productivity shocks $(\omega_{it}^L, \omega_{it}^H)$, and input prices (p_t^l, p_t^m).*

$$\min_{L_{it},M_{it}} p_t^l L_{it} + p_t^m M_{it} \quad s.t. \quad F_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big)\exp(\omega_{it}^H)\mathcal{E}_{it} \geqslant \bar{Y}_{it} \tag{2.3}$$

where $\mathcal{E}_{it} := \mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}]$.[8] Under cost minimization, the firm chooses its level

[6]Note that in this example, as $\rho_1 \to -\infty$ the production function approaches Leontief in materials and labor, so the model can approximate the Leontief production function.

[7]Homothetic separability is not satisfied by all parametric production functions. For example, the CES production function with nested capital and labor violates this property. Nadiri (1982) provides additional examples and discusses properties of non-homothetic production functions.

[8]Since ϵ_{it} is independent of $\mathcal{I}_{it}$, $\mathbb{E}[\exp(\epsilon_{it})|\mathcal{I}_{it}]$ becomes constant and does not affect the cost minimization problem of the firm.

of flexible inputs to minimize the cost of producing its profit-maximizing target output $\bar{Y}_{it}$, which can deviate from observed output Y_{it} due to the unanticipated shock ϵ_{it}. The firm determines $\bar{Y}_{it}$ by setting marginal revenue equal to marginal cost, so it depends on factors affecting marginal revenue, such as demand shocks (Doraszelski and Jaumandreu 2023; Ackerberg and De Loecker 2024). As I discuss later, this dependence imposes certain restrictions on product market competition.

Proposition 2.1.

(i) *Under Assumptions 2.1(i-ii) and 2.2, the optimal flexible input ratio, denoted by $\tilde{M}_{it} = M_{it}/L_{it}$, depends only on K_{it} and ω_{it}^L through an unknown function $r_t(\cdot)$:*

$$\tilde{M}_{it} \equiv r_t(K_{it}, \omega_{it}^L). \tag{2.4}$$

(ii) *Under Assumption 2.1(iii), $r_t(K_{it}, \omega_{it}^L)$ is strictly monotone in ω_{it}^L.*

The proof is provided in Appendix A. Part (i) of the proposition states that the firm's cost-minimizing flexible input ratio is a function of only K_{it} and labor-augmenting productivity ω_{it}^L. This result follows directly from the homotheticity assumption: the relative marginal product of materials and labor depends only on their ratio, not on their levels, and it is unaffected by the Hicks-neutral productivity ω_{it}^H. Note also that $\tilde{M}_{it}$ depends on the input price ratio, which is captured by the time subscript in $r_t(\cdot)$ as input prices are assumed constant across firms.[9]

Part (ii) establishes that $r_t(K_{it}, \omega_{it}^L)$ is strictly monotone in ω_{it}^L. Strict monotonicity implies that as labor-augmenting productivity ω_{it}^L increases, the materials-to-labor ratio consistently moves in a single direction. This direction is determined by the elasticity of substitution, as changes in ω_{it}^L alter the relative marginal products of materials and labor. Since these inputs are assumed to be either substitutes or complements everywhere, an increase in ω_{it}^L affects their ratio monotonically.[10]

To relate these results to parametric production functions, note that under CES, $r_t(K_{it}, \omega_{it}^L)$ has a log-linear form in ω_{it}^L: $\log(\tilde{M}_{it}) = \gamma + \sigma \log(p_t^{l/m}) + (\sigma - 1)\log(\omega_{it}^L)$, where $p_t^{l/m}$ is the input price ratio and γ is a constant that depends on the parameters (β, σ). The literature commonly uses this relationship to estimate labor-augmenting productivity, exploiting the linear separability of ω_{it}^L

[9]For $\tilde{M}_{it}$ function under time-varying input prices, see Equation (OA-1) in Online Appendix A.

[10]Another common assumption in production functions is Leontief, where inputs are perfect complements. In this case, $r_t(\cdot)$ becomes a multiplicative function of ω_{it}^L conditional on K_{it}.

in the input demand function under the CES form and variation in input prices (Doraszelski and Jaumandreu 2018; Raval 2019). Thus, this paper extends the CES framework to an arbitrary functional form (subject to Assumption 2.1) and shows that invertibility holds under more general conditions.[11,12]

2.2.2 Other Assumptions

The remaining assumptions extend the commonly used production function estimation assumptions to accommodate labor-augmenting technology.

Assumption 2.3 (First-Order Markov)**.** *Productivity shocks follow an exogenous joint first-order Markov process:* $P(\omega_{it}^L, \omega_{it}^H \mid \mathcal{I}_{it-1}) = P(\omega_{it}^L, \omega_{it}^H \mid \omega_{it-1}^L, \omega_{it-1}^H)$.

This assumption extends the standard first-order Markov property to a joint one to accommodate two-dimensional productivity. Importantly, it does not constrain the joint distribution of productivity shocks or their first-order dynamics, allowing productivity shocks to be arbitrarily correlated with one another.

Assumption 2.4 (Timing)**.** *The firm's capital stock evolves according to* $K_{it} = (1-\delta)K_{it-1} + I_{it-1}$, *where* I_{it-1} *denotes investment made by firm* i *during period* $t-1$ *and* δ *denotes the depreciation rate.*

Firms choose their capital one period ahead, before observing current productivity. Thus, K_{it} belongs to the firm's information set in period $t-1$, that is, $K_{it} \in \mathcal{I}_{it-1}$.

Assumption 2.5 (Monotonicity)**.** *Firms' materials demand is given by*

$$M_{it} = s_t(K_{it}, \omega_{it}^H, \omega_{it}^L), \tag{2.5}$$

where $s_t(K_{it}, \omega_{it}^H, \omega_{it}^L)$ *is an unknown function that is strictly increasing in* ω_{it}^H.

Introduced by Levinsohn and Petrin (2003), this assumption states that holding all else constant, more productive firms have higher materials demand. Moreover, the input demand function depends only on capital stock and productivity shocks.

[11] CES may not be restrictive for certain empirical questions—for example, when studying the elasticity of substitution—as it provides a first-order approximation to any production function with separability (Doraszelski and Jaumandreu 2018).

[12] Note that $r_t(\cdot)$ is not completely unknown, as it is derived from the production function (see Equation (A.5) in Appendix A). In my model, unlike in a parametric setting, the relationship between $r_t(\cdot)$ and the production function is complex, involving inverse functions. This complexity prevents the application of additional restrictions on $r_t(\cdot)$ that a parametric specification could allow. For example, under the CES assumption, the estimation of $r_t(\cdot)$ gives the elasticity of substitution parameter, whereas this is not possible in my specification.

To understand the underlying conditions under which this assumption holds, consider the firm's cost minimization problem in Equation (2.3). Under cost minimization, the firm's input demand depends on its profit-maximizing output level, $\bar{Y}_{it}$, which itself is influenced by demand shifters and the form of competition in the product market. Therefore, a natural question is under which product market assumptions the firm's profit-maximizing output $\bar{Y}_{it}$ depends solely on its own cost determinants $(K_{it}, \omega^H_{it}, \omega^L_{it})$, thereby ensuring that Assumption 2.5 holds.

I examine this question in Online Appendix B and show that firms' materials demand follows the form in Equation (2.5) under two conditions. The first condition requires the competition game to be symmetric (up to observables if one conditions on them). Symmetry implies that there is no unobserved firm-level heterogeneity in firms' residual demand; otherwise, the materials demand function would involve firm-specific unobserved demand shocks, violating the two-dimensional unobserved heterogeneity structure required for identification.

The second condition requires the competition game to be *aggregative* (Jensen 2018; Nocke and Schutz 2018). Even in symmetric competition games, a firm's output choice $\bar{Y}_{it}$ may depend on competitors' productivity shocks, violating the assumption of two-dimensional unobserved heterogeneity in material demand. Aggregative games, however, have the property that a firm's profit depends on its rivals' actions only through industry-level aggregate variables. By leveraging this property, I show that when the game is aggregative, a firm's profit-maximizing output depends solely on firm-specific cost factors (K_{it}, ω^L_{it}, ω^H_{it}) and industry-level aggregates common to all firms, represented by the time index t in materials demand function $s_t(\cdot)$. I further show that some standard imperfect competition models—symmetric Cournot, Bertrand competition with logit and CES demand systems, and monopolistic competition—can be formulated as aggregative games, leading to the materials demand function in Equation (2.5).[13] A key condition for these results is a weakly convex static cost function, which allows for non-constant marginal cost and is empirically supported by the estimates reported in Section 6.

The class of symmetric aggregative games is quite restrictive, as it excludes

[13]The application of aggregative game theory to imperfect competition models is used extensively in the literature (Caplin and Nalebuff 1991; Anderson et al. 2020). My analysis relies on these papers and extends some of their results to non-constant marginal costs.

many prominent demand models in industrial organization (IO), including random coefficient logit models (Berry et al. 1995). However, tools from aggregative game theory have proven useful in IO recently to reduce the dimension of many complex problems (Garrido 2022; Nocke and Whinston 2022; Caradonna et al. 2025). Incorporating imperfect competition into production function estimation remains an active area of research; see Ackerberg and De Loecker (2024) for a full treatment of various imperfect competition models in production function estimation.[14]

2.3 Invertibility: Expressing Productivity Shocks Using Inputs

Proposition 2.1 provides the sufficient conditions—monotonicity and scalar unobserved heterogeneity—to invert out ω_{it}^L using the flexible input ratio:

$$\omega_{it}^L = r_t^{-1}(K_{it}, \tilde{M}_{it}) \equiv \bar{r}_t(K_{it}, \tilde{M}_{it}). \tag{2.6}$$

Similarly, Assumption 2.5 provides a monotonicity condition for ω_{it}^H using the materials demand function in Equation (2.5). Inverting Equation (2.5) yields $\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \omega_{it}^L)$. Substituting for ω_{it}^L from Equation (2.6) gives:

$$\omega_{it}^H = s_t^{-1}(K_{it}, M_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it})) \equiv \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}). \tag{2.7}$$

Equations (2.6) and (2.7) demonstrate that, under the modeling assumptions and optimal firm behavior, unobserved productivity shocks can be written as unknown functions of inputs. Invertibility is a common feature in the production function literature, which uses observables, such as investments or materials, to control for unobserved productivity (Ackerberg et al. 2015). In the next section, I apply these invertibility results to develop a control variable approach to address endogeneity.

3 A Control Variable Approach to Production Functions

This section uses the Markov and timing assumptions to construct a control variable for each productivity shock. I build on the identification method proposed by Imbens and Newey (2009) for non-separable models with a scalar unobservable, which has recently been applied to production function estimation (Ackerberg et al. 2023; Pan 2024). I extend this control variable approach to accommodate

[14] Another key element of Assumption 2.5 is monotonicity of materials demand in ω_{it}^H. The relationship between input demand and productivity is complex because it depends on both the demand curve and the firm's cost function. Biondi (2022) provides a recent analysis of the relationship between factor demand and productivity under variable markups.

multi-dimensional productivity by leveraging two insights. First, the Markov and timing assumptions imply a statistical independence condition that is necessary to apply the Imbens and Newey (2009) approach. Second, the triangular structure of the input demand functions in Equations (2.4) and (2.5) enables the use of control variables in the presence of multidimensional productivity.

3.1 Control Variable for Labor-Augmenting Productivity

We can relate labor-augmenting productivity to lagged productivity as follows:

$$\omega_{it}^L = g_L(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1), \qquad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H \sim \mathrm{U}(0,1). \tag{3.1}$$

This representation of ω_{it}^L holds without loss of generality and follows from the Skorohod representation of random variables (Chernozhukov and Hansen 2008). Here, $g_L(\omega_{it-1}^L, \omega_{it-1}^H, \tau)$ represents the τ-th conditional quantile of ω_{it}^L given $(\omega_{it-1}^L, \omega_{it-1}^H)$, so u_{it}^1 can be interpreted as firm i's productivity rank among firms with the same lagged productivity. Another way to interpret u_{it}^1 is as the unanticipated *innovation* to ω_{it}^L. Unlike the standard definition of "innovation" to productivity in the production function literature (Ackerberg et al. 2015)—where the innovation is separable from and mean-independent of lagged productivity—u_{it}^1 is non-separable and fully independent of lagged productivity.

Recall from the previous section that $\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^L)$. By substituting ω_{it}^L from Equation (3.1) into this expression and using Equations (2.6-2.7), we obtain

$$\begin{aligned}
\tilde{M}_{it} &= r_t\big(K_{it}, g_L(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)\big), \\
&= r_t\big(K_{it}, g_L\big(\bar{r}_{t-1}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}_{t-1}(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1\big)\big), \\
&\equiv \tilde{r}_t\left(K_{it}, W_{it-1}, u_{it}^1\right),
\end{aligned} \tag{3.2}$$

for an unknown function $\tilde{r}_t(\cdot)$ and $W_{it-1} := (K_{it-1}, M_{it-1}, L_{it-1})$.[15] Note that $\tilde{M}_{it}$ is strictly monotone in u_{it}^1 because $r_t(\cdot)$ is strictly monotone in ω_{it}^L by Proposition 2.1, and $g_L(\cdot)$ is strictly monotone in u_{it}^1 by construction. Next, I establish the statistical independence of u_{it}^1 from other variables in Equation (3.2).

Lemma 3.1. *Under Assumptions 2.3 - 2.4, we have that* $u_{it}^1 \perp\!\!\!\perp (K_{it}, W_{it-1})$.

See Appendix A for the proof. This lemma shows that the Markov and timing assumptions provide the necessary independence condition for using the control

[15] Although $r_t(\cdot)$ depends on $\tilde{M}_{it-1}$, it is not included in W_{it-1} as it is a function of (M_{it-1}, L_{it-1}).

variable approach. The intuition behind this result is as follows: once we condition on $(\omega_{it-1}^L, \omega_{it-1}^H)$, the variables (K_{it}, W_{it-1}) provide no additional information about current productivity due to the timing and Markov assumptions. Since by construction Equation (3.1) implies that u_{it}^1 encapsulates all new information about current productivity, it follows that (K_{it}, W_{it-1}) must be independent of u_{it}^1.

We now have the two conditions needed to derive a control variable: monotonicity of $\tilde{r}_t(\cdot)$ in u_{it}^1 and independence of u_{it}^1 from (K_{it}, W_{it-1}). Given that the distribution of u_{it}^1 is already normalized to a uniform distribution in Equation (3.1), we can identify u_{it}^1 from the data using Equation (3.2) as follows:

$$u_{it}^1 = F_{\tilde{M}_{it}|K_{it},W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}), \tag{3.3}$$

where $F_{\tilde{M}_{it}|K_{it},W_{it-1}}$ denotes the CDF of $\tilde{M}_{it}$ conditional on (K_{it}, W_{it-1}).[16] The key insight is that if two firms, i and j, have the same capital stock and lagged inputs but differ in their materials-to-labor ratios, then they must differ only in their innovations to labor-augmenting productivity. Specifically, if $K_{it} = K_{jt}$ and $W_{it-1} = W_{jt-1}$, then $\tilde{M}_{it} > \tilde{M}_{jt}$ if and only if $u_{it}^1 > u_{jt}^1$. Thus, u_{it}^1 can be recovered from the firm's conditional rank in the flexible input ratio. This relationship allows us to express ω_{it}^L as a function of the control variable and lagged inputs:

$$\begin{aligned}\omega_{it}^L &= g_L(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) = g_L\left(\bar{r}_{t-1}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}_{t-1}(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}), u_{it}^1\right),\\ &\equiv c_{1t}(W_{it-1}, u_{it}^1), \end{aligned} \tag{3.4}$$

where $c_{1t}(\cdot)$ is an unknown function.

3.2 Control Variable for Hicks-Neutral Productivity

The control variable for ω_{it}^H can be derived similarly to that for ω_{it}^L. Writing the Skorohod representation of ω_{it}^H:

$$\omega_{it}^H = g_H(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \mathrm{U}(0,1). \tag{3.5}$$

By following the same steps used to obtain Equation (3.2), and using the monotonicity of materials in ω_{it}^H (Assumption 2.5), we can write materials demand as:

$$M_{it} \equiv \tilde{s}_t\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right), \tag{3.6}$$

[16]To simplify the exposition, I assume $\tilde{M}_{it}$ is strictly increasing in u_{it}^1. This is without loss of generality because u_{it}^1 only needs to be identified up to a monotonic transformation.

where $\tilde{s}_t(\cdot)$ is an unknown function. Note that $\tilde{s}_t\left(K_{it}, W_{it-1}, u_{it}^1, u_{it}^2\right)$ is strictly increasing in u_{it}^2 because $s_t(K_{it}, \omega_{it}^L, \omega_{it}^H)$ is strictly increasing in ω_{it}^H by Assumption 2.5, and $g_H(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2)$ is strictly increasing in u_{it}^2 by construction.

Lemma 3.2. *Under Assumptions 2.3 - 2.4, we have that* $u_{it}^2 \perp\!\!\!\perp (K_{it}, W_{it-1}, u_{it}^1)$.

See Appendix A for the proof. We can now use Equation (3.6) to identify u_{it}^2 as:

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1). \tag{3.7}$$

Using this result, we can express ω_{it}^H as follows:

$$\omega_{it}^H \equiv c_{2t}(W_{it-1}, u_{it}^1, u_{it}^2) \tag{3.8}$$

for an unknown function $c_{2t}(\cdot)$. Together, this result and Equation (3.4) imply that lagged inputs and two control variables can control for productivity shocks.

A related study by Ackerberg et al. (2023) also uses control variables in production function estimation. Their model has a single productivity shock in an unrestricted manner, whereas my approach incorporates two productivity shocks within a homothetic separability structure. Another difference lies in the construction of the control variables: Ackerberg et al. (2023) use lagged inputs, whereas I use the input demand functions. These approaches complement each other, as the plausibility of the underlying assumptions depends on the specific context.

4 Identification

This section examines the identification properties of the production function. I first establish a functional dependence problem that prevents the identification of individual output elasticities and labor-augmenting productivity using only inputs and output data. However, the sum of flexible input elasticities remains identifiable, and by combining this sum with revenue shares of inputs, the FOCs enable separate identification of labor and materials elasticities. I then consider a stronger form of homothetically separable production function and show that all elasticities and labor-augmenting productivity are identified under this form.

4.1 A Non-identification Result

The production function can be written in logarithmic form as:

$$y_{it} = f_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it},$$

where $y_{it} = \log(Y_{it})$ and $f_t(\cdot) = \log(F_t)$. Since $h_t(\cdot)$ is homogeneous in its second and third arguments, we can rewrite the production function as follows:

$$y_{it} = f_t\big(K_{it}, L_{it} h_t(K_{it}, \omega^L_{it}, \tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}. \tag{4.1}$$

This reformulation is useful because it isolates ω^L_{it} as a direct argument of $h_t(\cdot)$. By substituting $\omega^L_{it} = \bar{r}_t(K_{it}, \tilde{M}_{it})$ into Equation (4.1), we obtain:

$$y_{it} = f_t\Big(K_{it}, L_{it} h_t\big(K_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big)\Big) + \omega^H_{it} + \epsilon_{it}.$$

This representation of the production function reveals an identification problem.

Proposition 4.1. *Without further restrictions, $h_t(\cdot)$ cannot be identified from variations in $(Y_{it}, K_{it}, L_{it}, M_{it})$.*

Proof. For fixed values of $(K_{it}, \tilde{M}_{it})$, the second argument of the function $h_t(\cdot)$ is uniquely determined. Thus, the data provide no independent variation in $(K_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$ to trace out all dimensions of $h_t(\cdot)$, even if $\bar{r}_t(\cdot)$ is known. This implies that $h_t(\cdot)$ is not identified from variations in $(Y_{it}, K_{it}, L_{it}, M_{it})$.[17] □

This result reveals a fundamental identification problem, as many potential objects of interest, such as output elasticities, are functions of $h_t(\cdot)$. To see this, the output elasticities can be written as (suppressing function arguments):

$$\theta^K_{it} := (f_{t1} + f_{t2} h_{t1}) K_{it}, \qquad \theta^L_{it} := f_{t2} h_{t2} L_{it} \bar{r}_t(K_{it}, \tilde{M}_{it}), \qquad \theta^M_{it} := f_{t2} h_{t3} M_{it},$$

where f_{tk} denotes the derivative of f_t with respect to its $k-$th argument and θ^X_{it} denotes the output elasticity of input X. Observe that all the output elasticities depend on the derivatives of $h_t(\cdot)$, which is not identified from variations in inputs and output. However, if there is firm-level variation in input prices, then $\bar{r}_t(K_{it}, \tilde{M}_{it})$ becomes a function of the input price ratio as well, which could resolve the functional dependence problem. This case is analyzed in Online Appendix A.

To make progress toward the identification of elasticities, I define another function, $\bar{h}_t(K_{it}, \tilde{M}_{it}) \equiv h_t(K_{it}, \bar{r}_t(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$, and rewrite the production function as:

$$y_{it} = f_t\big(K_{it}, L_{it} \bar{h}_t(K_{it}, \tilde{M}_{it})\big) + \omega^H_{it} + \epsilon_{it}. \tag{4.2}$$

[17] Online Appendix C.1 illustrates this identification problem in a parametric setting using the CES production function. In that case, estimating a reduced form of the CES production function recovers a parameter corresponding to the sum of labor and materials elasticities.

This formulation provides a *reduced form* representation of the production function, where $\bar{h}_t(\cdot)$ captures the combined effects of ω_{it}^L and $\tilde{M}_{it}$ in production through the firm's input demand function $\bar{r}_t(K_{it}, \tilde{M}_{it})$. Even though this function differs from the structural function $h_t(\cdot)$, it can still be useful for identifying certain objects of interest. Thus, the remainder of this section focuses on what can be identified from reduced form functions $\bar{h}_t(\cdot)$ and $f_t(\cdot)$, while the next section examines the identification of these functions after specifying the moment conditions.

4.2 Identification of Output Elasticities

The functional dependence problem presented in Section 4.1 implies that output elasticities are not identified from variation in inputs and output alone. However, the assumption of cost minimization can provide additional identifying information through the relationship between the production function and the firm's flexible input choices. The FOCs derived from the cost minimization in Equation (2.3) are given by $F_{tX}(\cdot)\exp(\omega_{it}^H)\mathbb{E}[\exp(\epsilon_{it})|\mathcal{I}_{it}]\lambda_{it} = p_t^X$ for $X \in \{M, L\}$. Here, $F_{tX}(\cdot)$ denotes the derivative of $F_t(\cdot)$ with respect to input X, and λ_{it} represents the Lagrange multiplier. Multiplying both sides by $X_{it}/(Y_{it}p_{it})$ and rearranging yields:

$$\underbrace{\frac{F_{tX}\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big)X_{it}}{F_t\big(K_{it}, \omega_{it}^L L_{it}, M_{it}\big)}}_{\text{Elasticity}(\theta_{it}^X)} \frac{\mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}]\lambda_{it}}{\exp(\epsilon_{it})p_{it}} = \underbrace{\frac{X_{it}p_t^X}{Y_{it}p_{it}}}_{\text{Revenue Share of Input}(\alpha_{it}^X)}, \tag{4.3}$$

where p_{it} is the output price. Taking the ratio of this expression for labor ($X = L$) and materials ($X = M$) yields:

$$\theta_{it}^M/\theta_{it}^L = \alpha_{it}^M/\alpha_{it}^L. \tag{4.4}$$

Thus, the cost minimization assumption identifies the materials-to-labor elasticity ratio as the ratio of their revenue shares. Since revenue shares are observable in most datasets, this elasticity ratio can be directly calculated from data without any additional estimation.[18]

The use of FOCs in production function estimation has a long history, primarily within parametric frameworks (Kmenta 1967; Doraszelski and Jaumandreu 2013; Grieco et al. 2016). Gandhi et al. (2020) advance this literature by nonparametrically using the FOC with respect to materials input in a Hicks-neutral production

[18] Doraszelski and Jaumandreu (2019) similarly exploit revenue shares to identify elasticity ratios.

function with dynamic labor input. I instead restrict labor to be flexible, enabling me to use FOCs for two inputs to accommodate labor-augmenting productivity. This method of using FOCs of multiple flexible inputs has also been applied in the literature to estimate buyer markdowns (Morlacco 2020; Yeh et al. 2022).

4.2.1 Identification of the Sum of Materials and Labor Elasticities

This section shows the identification of the sum of labor and materials elasticities from the reduced-form representation of the production function in Equation (4.2).

Proposition 4.2. *The sum of labor and materials elasticities is identified as*

$$\theta_{it}^V := \theta_{it}^M + \theta_{it}^L = f_{t2}\big(K_{it}, L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it})\big) L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it}). \tag{4.5}$$

Proof. From Equation (4.1), the labor and materials elasticities can be written as:

$$\theta_{it}^M = f_{t2} h_{t3} M_{it}, \qquad \theta_{it}^L = f_{t2}(h_t - h_{t3}\tilde{M}_{it}) L_{it}.$$

Summing the expressions for θ_{it}^M and θ_{it}^L implies that the derivative component of $h_t(\cdot)$ cancels out, leaving the combined elasticity dependent on the level of $h_t(\cdot)$:

$$\theta_{it}^V = \theta_{it}^M + \theta_{it}^L = f_{t2} h_t L_{it} = f_{t2}\bar{h}_t L_{it}. \qquad \square$$

This proposition shows that identifying $f_t(\cdot)$ and $\bar{h}_t(\cdot)$ is sufficient for recovering the sum of flexible input elasticities. Notably, this identification does not rely on knowledge of either the structural function $h_t(\cdot)$ or the labor-augmenting productivity. By combining this result with the elasticity ratio derived from the FOCs in Equation (4.4), we can express the labor and materials elasticities as:

$$\theta_{it}^L = \theta_{it}^V \alpha_{it}^L / \alpha_{it}^V, \qquad \theta_{it}^M = \theta_{it}^V \alpha_{it}^M / \alpha_{it}^V, \tag{4.6}$$

where $\alpha_{it}^V = \alpha_{it}^L + \alpha_{it}^M$. Thus, labor and materials elasticities are separately identified. This result is especially important for markup estimation since flexible input elasticity is the key element in the markup formula, as I will discuss in Section 7.

4.2.2 Identification of Capital Elasticity and Other Objects

This section analyzes whether capital elasticity, the elasticity of substitution, and labor-augmenting productivity are identified and presents a negative result.

Proposition 4.3. *For the production function specified in Equation (2.2), labor-augmenting productivity, the output elasticity of capital, and the elasticity of substitution between inputs are not identified from* $(f_t, \bar{h}_t, \alpha_{it}^L, \alpha_{it}^M)$.

See Online Appendix D for the proof. The intuition underlying this result is that the FOCs reveal information primarily about the output elasticities of flexible inputs, which are insufficient to identify other aspects of the production function.

4.3 Identification under Further Restrictions

A potential solution to the non-identification results is to impose additional structure on the production function. In this section, I consider a more restrictive homothetic functional form:

$$y_{it} = f_t\big(K_{it}, h_t(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it}, \tag{4.7}$$

which I refer to as the "strongly homothetic" production function.[19] This model differs from the "weakly homothetic" production function presented in Equation (2.2) because the function $h_t(\cdot)$ no longer depends on K_{it}. Since this is a special case, Proposition 2.1 applies with $\omega_{it}^L = \bar{r}_t(\tilde{M}_{it})$. Substituting this into Equation (4.7) yields the following reduced form production function:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{4.8}$$

Since K_{it} appears only in $f_t(\cdot)$ and not in $h_t(\cdot)$, this model offers a more tractable functional form for identifying the capital elasticity and labor-augmenting productivity. The following proposition establishes the identification of these objects.

Proposition 4.4. *Suppose that the production function has the strongly homothetic separable form in Equation (4.7). Then, the capital elasticity is identified, and labor-augmenting productivity is identified up to scale from* $(f_t, \bar{h}_t, \alpha_{it}^L, \alpha_{it}^M)$*:*

$$\theta_{it}^K = f_{t1}(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it}))K_{it}, \qquad \log(\omega_{it}^L) = \int_{\underline{\tilde{M}}}^{\tilde{M}_{it}} b_t(\bar{M})d\bar{M} + a_t, \tag{4.9}$$

where $f_{tj}(\cdot)$ *denotes the derivative of* $f_t(\cdot)$ *with respect to its* j*-th argument,* a_t *is an arbitrary constant and* $\underline{\tilde{M}}$ *is the lower bound of the support of* $\tilde{M}_{it}$*.* $b_t(\cdot)$ *is a function defined in the proof, which depends only on* $(f_t, \bar{h}_t, \alpha_{it}^L, \alpha_{it}^M)$*.*

[19] This definition differs from the classical "strong separability" concept in production theory. For different separability definitions, see Chambers (1988). Here, we also use the same function $h_t(\cdot)$ in both strong and weak homothetic functions with a slight abuse of notation.

The proof is provided in Online Appendix D. Under strong homotheticity, θ_{it}^K is identified from the reduced form functions because ω_{it}^L is no longer a function of capital, removing the confounding effects of $\bar{r}_t(\cdot)$ in the reduced form function. Identification of ω_{it}^L relies on the idea that the flexible input elasticities are informative about $h_t(\cdot)$, which can be mapped back to ω_{it}^L as shown in the proof.

The next result shows that the elasticity of substitution remains unidentified.

Proposition 4.5. *Under the conditions of Proposition 4.4, the elasticity of substitution between labor and materials is not identified from* $(f_t, \bar{h}_t, \alpha_{it}^L, \alpha_{it}^M)$.

See Online Appendix D for the proof. The FOCs provide information only about the first derivatives of the production function, whereas the elasticity of substitution depends on the second derivatives. Thus, we can identify the output elasticities but not the elasticity of substitution.[20]

An important implication of using FOCs for identification is that elasticities can be identified only at values observed in the data, specifically within the set $\{(\omega_{it}^L, M_{it}, L_{it}) : \omega_{it}^L = \bar{r}_t(\tilde{M}_{it})\}$. Although this restriction limits certain counterfactual analyses—such as the impact of altering an input on output while holding ω_{it}^L constant—it nonetheless allows for the identification of elasticities and productivity for the firms in the data, which are sufficient for most empirical applications.

5 Empirical Model and Data

This section presents the empirical model, outlines the estimation procedure, and describes the data used for empirical applications.

5.1 Empirical Model

The purpose of the empirical model is to estimate output elasticities and markups. Because the capital elasticity is not identified under weak homotheticity, and to reduce data requirements, I use the strongly homothetic production function in the empirical model. The resulting reduced form production function is given by:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{5.1}$$

[20] This finding relates to Diamond et al. (1978)'s impossibility theorem, which states that without exogenous variation in input prices, the elasticity of substitution cannot be identified using time-series data. My paper extends this result to a setting with firm-level data. Under the heterogeneous input price extension discussed in Online Appendix A, however, the elasticity of substitution can be identified if input prices are exogenous.

To control for Hicks-neutral productivity, I use the control function from Equation (3.8), $\omega_{it}^H = c_{2t}\left(W_{it-1}, u_{it}^1, u_{it}^2\right)$. Substituting this into Equation (5.1) yields:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}\left(W_{it-1}, u_{it}^1, u_{it}^2\right) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0.$$

Because all the right-hand side variables are orthogonal to the error term ϵ_{it}, we can estimate this equation by minimizing the sum of squared residuals. The modeling assumptions, however, provide additional restrictions that we can use to augment the moment conditions. Specifically, since capital input is predetermined, it is orthogonal to productivity innovations ξ_{it} from the first-order Markov process:

$$\omega_{it}^H \equiv \bar{c}_{3t}(\omega_{it-1}^L, \omega_{it-1}^H) + \xi_{it}, \qquad \mathbb{E}[\xi_{it} \mid \mathcal{I}_{it-1}] = 0. \tag{5.2}$$

where $\bar{c}_{3t}(\cdot)$ is an unknown function and ξ_{it} denotes the innovation to Hicks-neutral productivity. Unlike u_{it}^2 in Equation (3.5), this innovation term is mean independent of $(\omega_{it-1}^H, \omega_{it-1}^L)$ and separable, instead of being fully independent and non-separable. This characterization of ω_{it}^H has been frequently used in the production function literature for constructing moments (Ackerberg et al. 2015).

Since $(\omega_{it-1}^L, \omega_{it-1}^H)$ can be expressed as a function of W_{it-1}, Equation (5.2) can be written as $\omega_{it}^H \equiv c_{3t}(W_{it-1}) + \xi_{it}$, which yields another estimating equation:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{3t}(W_{it-1}) + \xi_{it} + \epsilon_{it}.$$

The error term, $\xi_{it} + \epsilon_{it}$, is orthogonal to the firm's information set at time $t-1$, which includes K_{it}, so $\mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}, W_{it-1}] = 0$. I now summarize the estimation problem by combining the moment restrictions from two estimating equations:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it},\ \mathbb{E}[\epsilon_{it} \mid W_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0 \tag{5.3}$$

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{3t}(W_{it-1}) + \xi_{it} + \epsilon_{it},\ \mathbb{E}[\xi_{it} + \epsilon_{it} \mid K_{it}, W_{it-1}] = 0. \tag{5.4}$$

The literature typically uses similar moment conditions in two stages when estimating parametric forms: first, identifying the elasticity of variable inputs, and then the capital elasticity (Ackerberg et al. 2015). This two-step method is infeasible in a non-separable production function. Thus, I combine all moment conditions into a single objective function, following the approach of Wooldridge (2009):

$$J(\theta) = \frac{1}{TN}\sum_{i,t}\widehat{\epsilon}_{1it}(\theta)^2 + \sum_{j=1}^{J}\left(\frac{1}{TN}\sum_{i,t} z_j(K_{it}, W_{it-1})[\widehat{\xi}_{it}(\theta) + \widehat{\epsilon}_{2it}(\theta)]\right)^2, \tag{5.5}$$

where $\theta = \{f_t, \bar{h}_t, c_{2t}, c_{3t}\}$ collects all unknown functions to be estimated. For each candidate θ, $\widehat{\epsilon}_{1it}(\theta)$ denotes the estimates of ϵ_{it} from Equation (5.3), while $\widehat{\epsilon}_{2it}(\theta)$ and $\widehat{\xi}_{it}(\theta)$ denote the estimates of ϵ_{it} and ξ_{it} from Equation (5.4). Finally, $\{z_j(K_{it}, W_{it-1})\}_{j=1}^{J}$ denote a set of J instruments.

The identification of output elasticities requires identifying the functions (f_t, $\bar{h}_t$) through the moment restrictions in Equations (5.3) and (5.4). I next show the identification of these functions. Define $g_t(K_{it}, M_{it}, W_{it-1}, u_{it}^1)$ as the conditional distribution function $F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$ and let $g_{tj}(\cdot)$ denote the derivative of $g_t(\cdot)$ with respect to its j-th arguments. The following proposition establishes that moment conditions in Equation (5.3) identify (f_t, $\bar{h}_t$).

Proposition 5.1. *Suppose that (i) $(f_t, \bar{h}_t, c_{2t}, g_t)$ are twice continuously differentiable and have non-zero derivatives almost everywhere, (ii) the joint distribution function of $(K_{it}, \tilde{M}_{it}, M_{it}, W_{it-1})$ is absolutely continuous with positive density everywhere on its support, (iii) $Var\big(g_{t1}(K_{it}, M_{it}, W_{it-1}, u_{it}^1)/g_{t2}(K_{it}, M_{it}, W_{it-1}, u_{it}^1) \mid K_{it}, M_{it}, u_{it}^1\big) > 0$. Then $\bar{h}_t(\cdot)$ is identified up to a scale, and $f_t(\cdot)$ is identified up to a constant by the moment conditions in Equation (5.3).*

See Appendix B for the proof.[21,22] Conditions (i) and (ii) are regularity conditions that ensure sufficient variation in the data. Condition (iii) rules out special cases that create functional dependence, a concept known as generic identification in the literature (Lewbel 2019). Since the moments in Equation (5.3) identify the target functions, the moment restrictions in Equation (5.4) provide efficiency gains.[23]

5.2 Estimation Procedure

This section outlines the estimation procedure, while the technical details are provided in Online Appendix E. I estimate separate production functions for each industry, based on the classifications provided in Table 1. Given the relatively small sample sizes in many industries, estimating the production function using a single year of data leads to noisy estimates. As a result, following De Loecker et al. (2020), I adopt a rolling window sample strategy, using an eight-year window for the Compustat data and a three-year window for the other datasets.

[21] Online Appendix D.1 presents the corresponding identification results for the weakly homothetic production functions given in Equation (2.2).

[22] Even though $\bar{h}_t(\cdot)$ is identified up to scale, θ_{it}^V is uniquely identified (see Proposition 4.2).

[23] Ideally, one would analyze the identification properties of the moment restrictions in Equations (5.3) and (5.4) jointly. I focus on Equation (5.3) due to the complexity of a joint analysis.

The estimation proceeds in two stages. In the first stage, I identify the control variables by estimating the CDF in Equation (3.7) through partitioning the support of M_{it} into 500 equally sized grids and fitting a logit model with third-degree polynomials at each grid point.[24] In the second stage, I estimate the production function using polynomial approximations. Specifically, I first approximate the logarithm of $\bar{h}_t(\cdot)$ with a third-degree polynomial:

$$\log(\widehat{\bar{h}}_t(\tilde{M}_{it})) = a_{1t} + a_{2t}\tilde{m}_{it} + a_{3t}\tilde{m}_{it}^2 + a_{4t}\tilde{m}_{it}^3, \tag{5.6}$$

where $\{a_{jt}\}_{j=1}^4$ are the parameters of the polynomial approximation and $\tilde{m}_{it} = \log(\tilde{M}_{it})$. I define $V_{it} := L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})$ and approximate the production function as:

$$\begin{aligned}\widehat{f}_t(K_{it}, V_{it}) = b_{1t} &+ b_{2t}k_{it} + b_{3t}k_{it}^2 + b_{4t}k_{it}^3 + b_{5t}v_{it} + b_{6t}v_{it}^2 + b_{7t}v_{it}^3 \\ &+ b_{8t}k_{it}^2 v_{it} + b_{9t}k_{it}v_{it}^2 + b_{10t}k_{it}v_{it},\end{aligned} \tag{5.7}$$

where $\{b_{jt}\}_{j=1}^{10}$ are the polynomial parameters.[25] I approximate $c_{2t}(\cdot)$ and $c_{3t}(\cdot)$ in the objective function similarly using third-degree polynomials.

Estimating $c_{2t}(\cdot)$ and $c_{3t}(\cdot)$ is computationally simple, as they can be partialled out for a given $(\widehat{f}_t, \widehat{\bar{h}}_t)$. Thus, the estimation reduces to finding the function $\widehat{f}_t(\cdot)$ and $\widehat{\bar{h}}_t(\cdot)$ that minimize the objective function given in Equation (5.5). After obtaining these estimates, I compute the output elasticities according to Equations (4.5), (4.6), and (4.9). For standard errors, I implement a bootstrap procedure with 250 replications, treating firms as independent observations in each resample.

5.3 Datasets

I use panel data from manufacturing industries in Chile, Colombia, India, Turkey, and the US. Table 1 presents descriptive statistics for each dataset, including the sample period, the number of industries, and the number of firms per year. Further details and additional descriptive statistics are provided in Online Appendix F.

5.3.1 Chile, Colombia, India, and Turkey

For Chile, Colombia, India, and Turkey, the data come from plant-level production information collected through manufacturing censuses. The Chilean dataset covers

[24] Under the strongly homothetic production function in Equation (4.7), u_{it}^1 corresponds to the CDF of $\tilde{M}_{it}$, so it does not need to be estimated in the first stage conditional on other variables.

[25] For the US, the third-degree polynomials give unstable estimates due to small sample size, especially in the early period. To address this, I use second-degree polynomials for the US.

Table 1: Descriptive Statistics on Datasets

	US	Chile	Colombia	India	Turkey
Sample Period	1961-2018	1979-96	1978-91	1998-2014	1983-2000
Num. of Industries	3	5	9	5	8
Industry Level	2-dig NAICS	3-dig SIC	3-dig SIC	3-dig NIC	3-dig SIC
Num. of Obs/Year	1247	2115	3918	2837	4997

Notes: This table provides descriptive statistics for each dataset. The number of industries reflects those included in the empirical estimation. Sample periods vary based on data availability.

plants with more than 10 employees for the period 1979–1996. Similarly, the Colombian data include all plants with more than 10 employees from 1981 to 1991. The Turkish dataset is based on the Annual Surveys of Manufacturing Industries and covers all establishments with 10 or more employees from 1983 to 2000. Finally, the source for Indian data is the Annual Survey of Industries, which includes plants with 100 or more employees from 1998 to 2014.

From each dataset, I construct input and output measures to estimate the production functions. Materials input is calculated by deflating materials cost with the appropriate deflators. Labor is measured by either the number of worker-days or the number of workers. Capital is derived using the perpetual inventory method or by deflating its book values. See Online Appendix F.6 for further details.

5.3.2 US

The Compustat sample comprises all publicly traded US manufacturing firms from 1961 to 2018. It includes data from financial statements, such as sales, total input expenditures, and number of employees. From these variables, I derive capital, labor, materials, and output measures. Output is measured as net deflated sales, and labor is measured by the number of employees. Because Compustat does not separately report expenditures on materials, I follow Keller and Yeaple (2009) and estimate them as the difference between "cost of goods sold and administrative and selling expenses" and the sum of depreciation and labor costs.

While the Compustat dataset is derived from financial accounting data rather than manufacturing censuses, it offers broader coverage across industries and over time. This wider coverage makes it particularly useful for studying changes in markups and has contributed to the growing evidence on the rise of market power in the US (Basu 2019; De Loecker et al. 2020).

5.3.3 Discussion of the Datasets

An important limitation of these datasets is that I mainly observe revenues and input expenditures rather than physical quantities, which could lead to biases if prices reflect firm-level demand heterogeneity (Doraszelski and Jaumandreu 2019, 2023; Bond et al. 2021). While some recent datasets include physical output and firm-level price indexes, I use several widely employed production datasets to provide systematic evidence across different empirical contexts. However, as a robustness check in Section 9, I use the quantity-based output measures and plant-level input prices from seven Indian industries that produce relatively homogeneous products. This analysis suggests that the paper's main results—comparing estimates between the labor-augmenting production function and alternative specifications—are robust to using quantities and controlling for input prices.

6 Empirical Results: Production Function

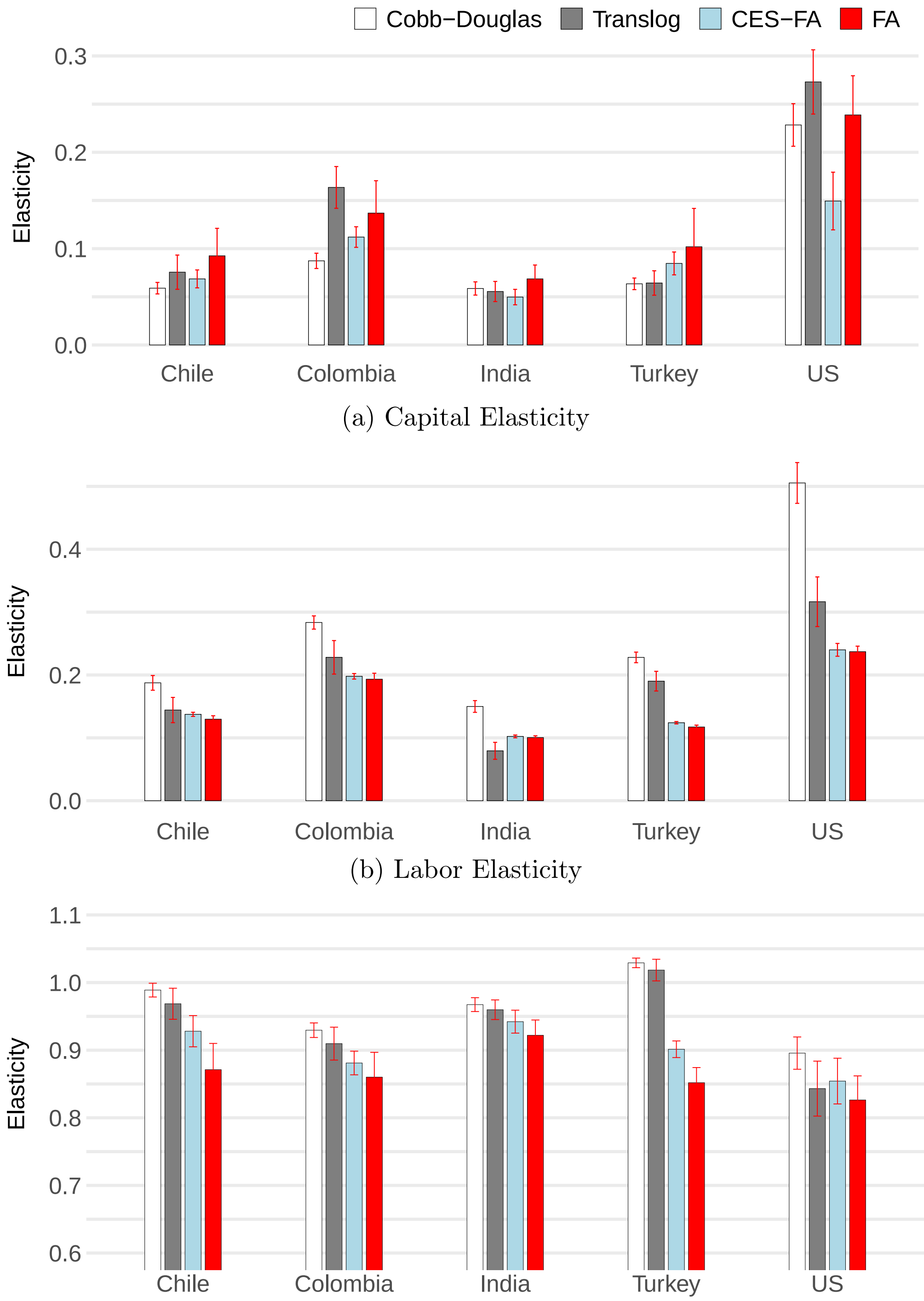
This section presents the production function estimates and emphasizes two main findings: (i) output elasticities vary systematically depending on whether labor-augmenting productivity is included in the specification, and (ii) there is substantial heterogeneity in output elasticities across firms, with larger firms being more capital-intensive than smaller firms.

6.1 Output Elasticities

I estimate output elasticities using four specifications with varying degrees of flexibility: (i) Cobb-Douglas (CD), (ii) Translog with Hicks-neutral productivity (TR), (iii) CES with both Hicks-neutral and labor-augmenting productivity (CES-FA), and (iv) the strongly homothetic production function with Hicks-neutral and labor-augmenting productivity (FA) introduced in Section 5.[26] Comparing these models allows me to evaluate two main features of the FA specification: its flexible functional form and its unobserved heterogeneity introduced through labor-augmenting productivity. In particular, because CES-FA embeds labor-

[26] Since Ackerberg et al. (2015) show that gross production functions with two flexible inputs are not identified using proxy variables, I estimate CD and TR using the Blundell and Bond (2000) dynamic panel method. I estimate CES-FA using procedures described in Section 5.2 after imposing the CES functional form assumption. The details of these estimation procedures are given in Online Appendix C.2 and Online Appendix E.2.

Figure 1: Average Elasticity Estimates Across Production Function Models



(a) Capital Elasticity

(b) Labor Elasticity

(c) Variable Input Elasticity

Notes: Comparison of sales-weighted average elasticities obtained from (i) Cobb-Douglas (CD), (ii) Translog with Hicks-neutral productivity (TR), (iii) CES with Hicks-neutral and labor-augmenting productivity (CES-FA), and (iv) the strongly homothetic production function with Hicks-neutral and labor-augmenting productivity (FA). For each year and industry, sales-weighted averages are calculated, and then simple averages of these industry-year estimates are taken over the years. The error bars indicate 95% confidence intervals calculated using bootstrap (250 resamples).

augmenting productivity within a parametric structure, the comparison between FA and CES-FA isolates the contribution of functional form flexibility. Conversely, TR offers functional form flexibility without labor-augmenting productivity, so comparing FA with TR isolates the role of labor-augmenting productivity.[27]

Figure 1 displays the sales-weighted average output elasticities for capital, labor, and variable input in each country.[28] Panel (a) shows that the CD specification consistently estimates lower capital elasticities than the FA model, with the bias ranging from -4.4% (US) to -37.7% (Turkey). When examining more flexible specifications, we observe that TR reduces this bias by an average of 57.1%, whereas CES-FA yields capital elasticities close to those obtained from CD. Together, these results indicate that functional form flexibility has a greater impact on capital elasticity estimation than allowing for labor-augmenting productivity.

Panel (b) reveals a contrasting pattern for labor elasticities: the CD specification systematically yields *higher* estimates than the FA model, indicating an upward bias ranging from 46.0% in Chile to 113.2% in Turkey. Unlike the capital elasticity results, CES-FA proves more effective than TR at reducing these biases (by 95.8% on average), suggesting that labor-augmenting productivity plays a greater role than functional form flexibility in estimating labor elasticity.

Moving to variable input elasticity in Panel (c), the estimates consistently decrease as the specification becomes more flexible—from CD to FA—in every country. Although both TR and CES-FA mitigate the bias present in the CD estimates, neither specification eliminates it entirely. The findings therefore point to roles for both functional form flexibility and unobserved heterogeneity in estimating variable input elasticity. Because variable input elasticity is central to markup estimation, this evidence also implies that incorporating labor-augmenting productivity could affect markup estimates, a topic I explore in the next section.

While this section focused on average elasticities, it is also important to examine the estimates in greater detail to evaluate their validity. For this purpose, Table OA-1 in Online Appendix H reports average elasticities for the three largest

[27] To be precise, CES-FA also adds functional form flexibility relative to CD by allowing for non-unitary elasticity of substitution. I choose the CES as the parametric form with labor-augmenting productivity because in the CD model, all productivity shocks are factor-neutral (Raval 2011).

[28] In Online Appendix H, Figure OA-1 reports materials elasticity and returns to scale estimates, and Figure OA-3 reports standard errors for differences between FA and other estimates.

industries in each country, while Supplementary Appendix B illustrates the distributions of first and second derivatives as well as the shapes of the estimated production functions.[29] The results align with the principles of production theory: capital elasticities are seldom negative (4.5%), first derivatives are predominantly positive (97.5%), and second derivatives are generally negative (89.4%), confirming the concavity of the estimated production functions (Shephard 1953).[30]

In summary, this section demonstrates how different sources of flexibility in the production function affect output elasticity estimates. For capital elasticity, functional form flexibility matters more than labor-augmenting technology, whereas for labor elasticity, labor-augmenting technology has a greater impact. Both types of flexibility, however, are important for estimating variable input elasticity.

6.2 Heterogeneity in Output Elasticities

This section explores firm-level heterogeneity in output elasticities and its relationship with firm size. While extensive research has shown substantial variation in various firm outcomes such as productivity, labor share, and management practices (Syverson 2011; Van Reenen 2018; Kehrig and Vincent 2021), comparatively little is known about heterogeneity in production technology.

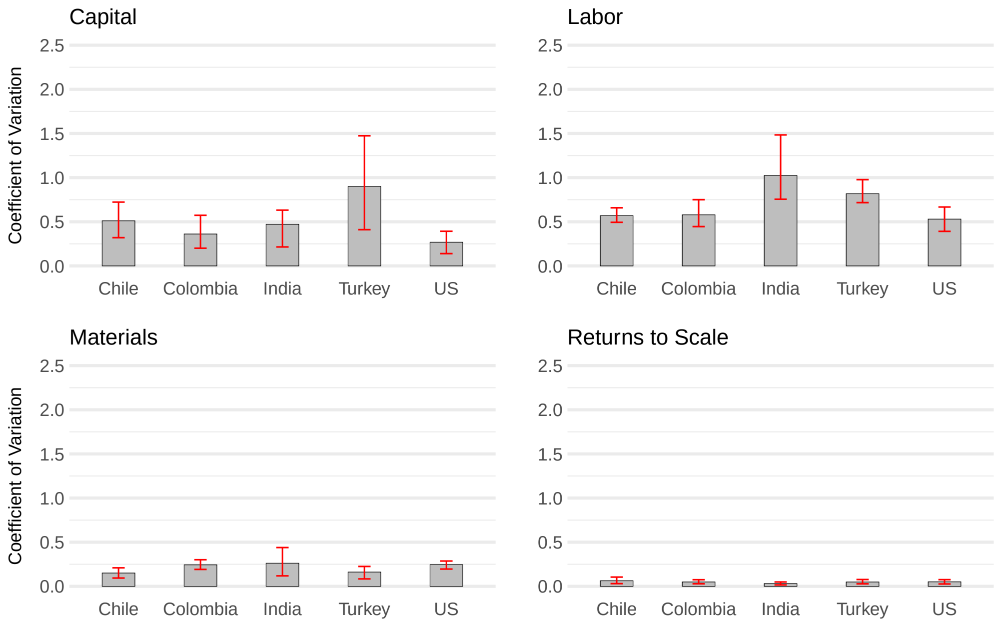
To quantify this heterogeneity, I calculate the coefficient of variation (CV) of output elasticities within each industry-year and plot the average CV in each country along with the 10th–90th percentile range in Figure 2. Elasticities vary substantially across firms in all countries, with CV estimates ranging from 0.25 to 1.01. Labor elasticity shows the greatest dispersion, consistent with evidence of large heterogeneity in labor's revenue share (Autor et al. 2020). The wide interdecile ranges confirm that these patterns are pervasive rather than driven by a handful of industries. Finally, returns to scale estimates show only modest variation—a reasonable finding since large heterogeneity in this parameter would imply some firms operate with implausibly strong scale economies.[31]

[29] The FA model generates reasonable output elasticities that align with previous literature: materials have the highest elasticity, ranging from 0.50 to 0.67 across industries and countries. The average labor and capital elasticities range from 0.22 to 0.52 and 0.04 to 0.16, respectively. The returns to scale estimates range from 0.93 to 1.1, indicating that firms, on average, operate close to constant returns to scale.

[30] The small number of negative elasticities can be partly explained by estimation error, as elasticities are not consistently estimated at the firm-level.

[31] The heterogeneity in output elasticities can arise from two sources: (i) observed heterogeneity

Figure 2: Average Coefficient of Variation Estimates of Output Elasticities



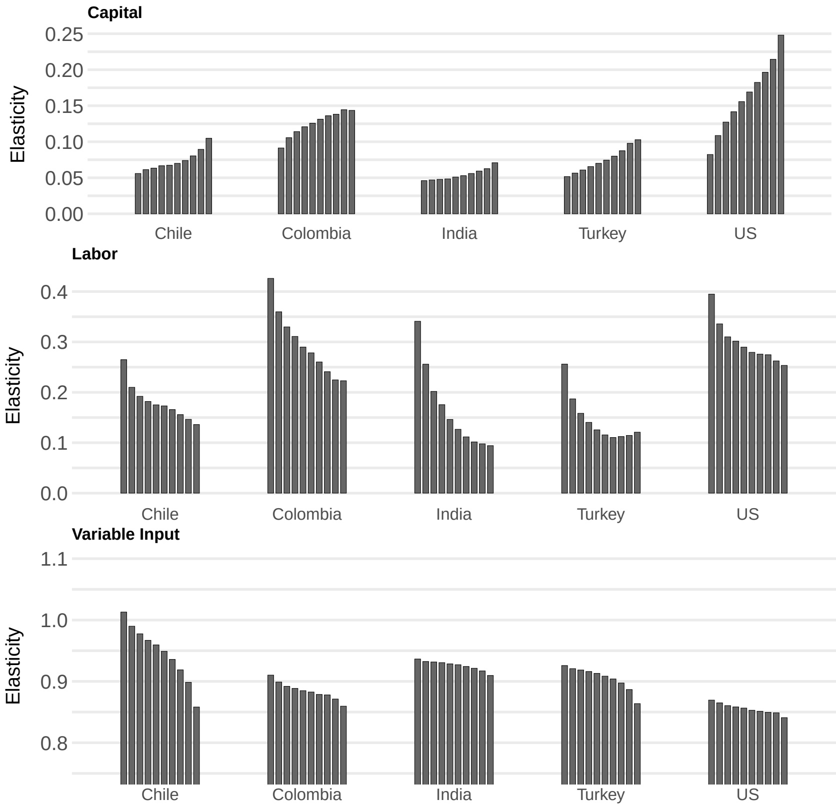
Notes: This figure shows the means (gray) and 10th and 90th percentiles (red) of the distribution of the average coefficient of variation across industries and years for different output elasticities.

The heterogeneity in output elasticities complements the well-documented dispersion in other firm-level outcomes. Yet another key question is whether this heterogeneity is correlated with observable firm characteristics. Figure 3 addresses this question by plotting average capital, labor, and variable input elasticities across firm-size deciles. Consistently across all countries, larger firms exhibit higher capital elasticities but lower labor and variable input elasticities. The differences are notable: firms in the largest-decile are about twice as capital-intensive as those in the smallest-decile, whereas the smallest-decile firms are roughly twice as labor-intensive as those in the largest. Overall, this section highlights the presence of substantial variation in output elasticities, which potentially reflects both observed and unobserved heterogeneity in production technologies across firms.[32]

due to different levels of inputs and (ii) unobserved heterogeneity due to labor-augmenting productivity. This analysis does not distinguish between these sources.

[32]For comparison, Figures SA-1 and SA-2 in Supplementary Appendix B report the corresponding results under the CES-FA and TR specifications (by construction, CD does not allow heterogeneous elasticities). Although these production functions yield some heterogeneity for certain inputs and countries, they do not generate a consistent pattern across inputs and countries.

Figure 3: Average Output Elasticity Estimates by Firm Size Deciles



Notes: This figure displays average output elasticity estimates across firm size deciles. For each country, I first estimate the average elasticity within each decile for each industry-year pair, then average these industry-year estimates within each country.

7 How Production Function Specification Affects Markups

Building on Hall (1988), De Loecker (2011) and De Loecker and Warzynski (2012) develop a production-based method for estimating markups under two assumptions: (i) firms minimize their static costs with respect to at least one flexible input, and (ii) they are price takers in the input market for that flexible input. Under these conditions, a firm's markup equals the output elasticity of a flexible input divided by that input's cost share in total revenue, $\mu_{it} := \theta_{it}^V / \alpha_{it}^V$.[33] Because the revenue shares of flexible inputs are typically observable in the data, a flexible input's output elasticity is sufficient to recover markups. This production-based approach to estimating markups has been widely adopted in recent studies of market power (Shapiro and Yurukoglu 2025; Miller 2025).

Because output elasticity enters directly into the markup formula, any bias

[33] As De Loecker and Warzynski (2012) note, the revenue share is with respect to planned output, which requires correcting for $\exp(\epsilon_{it})$. I implement this correction when estimating markups.

in its estimation translates into biased markups. This makes markup estimates sensitive to the choice of production function specification. Indeed, Van Biesebroeck (2008) finds that elasticity estimates can vary significantly across different estimation methods.[34] Motivated by this observation, this section analytically examines how biases in output elasticities impact the estimates of aggregate markup, followed by empirical evidence presented in the next section.

7.1 Conflicting Markup Estimates from Different Flexible Inputs

If firms minimize production costs with respect to two flexible inputs, the markup estimates derived from each input should coincide up to estimation error. However, empirical studies that rely on Hicks-neutral production functions often report systematic differences between markup measures based on different inputs (Doraszelski and Jaumandreu 2019; Raval 2023). These studies suggest that certain misspecifications in the underlying production model generate discrepancies in markup values, and they point to a lack of factor-augmenting productivity.[35]

A notable feature of the labor-augmenting production function and this paper's identification strategy is that markup estimates based on labor and materials are identical. This equivalence follows directly from the use of FOCs (Equation (4.4)) in estimation, which links the ratio of revenue shares to the ratio of elasticities:

$$\frac{\theta^L_{it}}{\theta^M_{it}} = \frac{\alpha^L_{it}}{\alpha^M_{it}} \quad \Longrightarrow \quad \mu^L_{it} = \frac{\theta^L_{it}}{\alpha^L_{it}} = \frac{\theta^M_{it}}{\alpha^M_{it}} = \mu^M_{it}\,, \tag{7.1}$$

where μ^L_{it} and μ^M_{it} denote labor- and materials-based markups, respectively. The inclusion of labor-augmenting productivity is key for achieving this equivalence. Without labor-augmenting productivity, the identity in Equation (7.1) fails to hold empirically (as confirmed by Figure OA-2 in Online Appendix H), indicating a potential misspecification. Labor-augmenting productivity introduces an additional dimension of unobserved heterogeneity. This in turn leads to exactly-identified output elasticities (see Section 4.2), resulting in an internally consistent model

[34] This is in contrast to productivity estimates, which are found to be robust to production function specification (Van Beveren 2012; Blackwood et al. 2021).

[35] Raval (2023) tests the production function approach by comparing markup estimates from two flexible inputs. His analysis reveals that labor-based and materials-based markup measures are negatively correlated in the cross-section and time series. He then examines possible mechanisms for this discrepancy, including production function heterogeneity, labor adjustment costs, and measurement error. He concludes that the most plausible explanation is the inability of the standard production functions to account for heterogeneity in production technology.

and identical markup estimates from labor and materials.

While conflicting markup estimates can also come from other misspecifications (e.g., adjustment costs or input market power), factor-augmenting productivity provides a theoretically grounded solution, particularly when the main concern is insufficient unobserved heterogeneity in the production function (Raval 2023).[36]

7.2 Decomposing Markups: The Role of Production Functions

This section performs a decomposition analysis to investigate how biases in output elasticities influence aggregate markup estimates. The analysis identifies two distinct channels: (i) bias in the average output elasticity and (ii) bias in the correlation between output elasticities and firm size. To illustrate these channels, I apply the Olley and Pakes (1996) decomposition (OP) to the aggregate log-markup, defined as $\tilde{\mu}_t = \sum_i w_{it} \log(\mu_{it}) = \sum_i w_{it} \log(\theta_{it}) - \sum_i w_{it} \log(\alpha_{it})$, where w_{it} denotes the aggregation weight, typically a measure of firm size. The OP decomposition separates a weighted average into its unweighted mean and the covariance between the weights and the variable of interest, resulting in the following decomposition:

$$\tilde{\mu}_t = \underbrace{\underbrace{\mathbb{E}[\log(\theta_{it})]}_{\text{Avg. Elas (1)}} + \underbrace{\mathrm{Cov}(w_{it}, \log(\theta_{it}))}_{\text{Heterogeneity in Elasticity (2)}}}_{\text{Estimation}} - \underbrace{\underbrace{\mathbb{E}[\log(\alpha_{it})]}_{\text{Avg. Share (3)}} - \underbrace{\mathrm{Cov}(w_{it}, \log(\alpha_{it}))}_{\text{Heterogeneity in Shares (4)}}}_{\text{Data}} \tag{7.2}$$

The aggregate log markup breaks down into four distinct components: two arising from output elasticities and two from revenue shares. This structure cleanly separates the estimated elements (output elasticities) from the observable quantities (revenue shares), enabling us to trace how biases in output elasticities affect aggregate markups.

Bias from the Average Output Elasticity. The first component of the decomposition is the mean output elasticity. Misspecifications in the production function can bias this term, directly affecting the aggregate markup. The elasticity estimates in Section 6.1 indicate that Hicks-neutral production functions tend to overestimate the flexible input elasticity. Accordingly, we expect this component of the bias to be positive.

[36] An important limitation of my approach is that it requires two flexible inputs as opposed to only one. This is particularly limiting if some inputs are subject to adjustment costs.

Bias from Heterogeneity in Production Technology. The second component reflects the covariance between firm size and the flexible input elasticity. Bias in this component occurs when elasticities vary systematically with firm size, yet the estimation fails to capture this relationship. The estimates presented in Section 6.2 indicate an inverse relationship between flexible input elasticity and firm size, implying that neglecting this pattern likely introduces a positive bias. Therefore, the decomposition analysis indicates that both sources of bias in the aggregate markup are likely positive and do not offset each other. Finally, these biases can influence not only the level of markups but also their trends—for example, if large firms increase their capital intensity over time and the estimation approach fails to account for this shift in production technology.

8 Empirical Results: Markups

This section compares markup estimates across alternative production function specifications. I first analyze how these specifications influence markup levels and then examine markup trends within the US manufacturing sector.

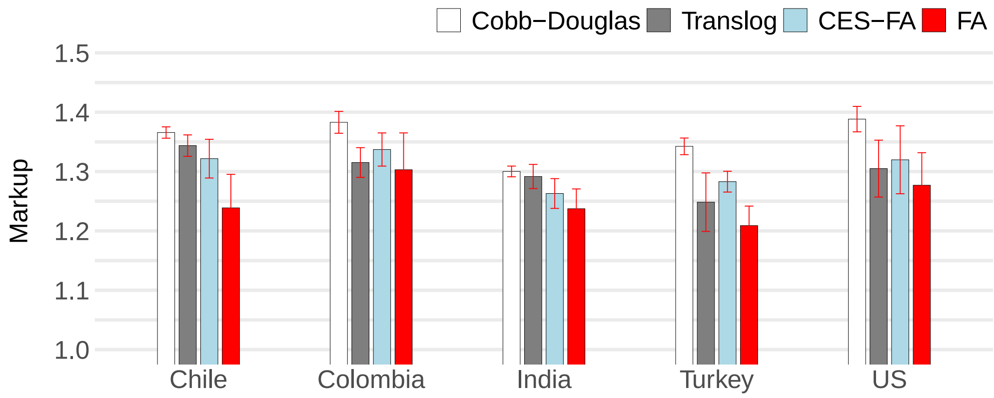
When estimating markups under Hicks-neutral production functions, I use a specification that includes capital and a single flexible input. I adopt this approach for two reasons. First, I observe that markup estimates derived from Hicks-neutral production functions using labor and materials differ markedly between the two inputs—often falling below one—as shown in Figure OA-2 and consistent with findings by Raval (2023). Second, the single flexible input specification facilitates comparisons with recent markup studies, which often use a single flexible input.[37]

8.1 Markups Comparison: Level

I compute sales-weighted markups for each country by averaging firm-level markup estimates derived from the output elasticities presented in Section 6.1. Figure 4 displays the resulting aggregate markups for four production functions—CD, TR,

[37] See, for example, De Loecker et al. (2020) and Autor et al. (2020). To give more details of the estimation, in the Cobb-Douglas case, I estimate the production function as $y_{it} = \beta_{kt} k_{it} + \beta_{vt} vr_{it} + \omega^H_{it} + \epsilon_{it}$, where vr_{it} combines labor and materials into a single flexible input. Markups can then be calculated using the estimated flexible input elasticity, $\mu_{it} = \beta_{vt}/\alpha^V_{it}$. Note that in this estimation, the parameters are allowed to change over time, enabling the CD model to capture shifts in the aggregate production technology. Thus, when comparing more flexible specifications with the CD model, differences in estimates should be interpreted as reflecting the additional flexibility offered by those specifications relative to the time-varying CD parameters.

Figure 4: Average Markup Estimates Across Production Function Models



Notes: Comparison of sales-weighted markups estimated using four production functions. Sales-weighted averages are calculated for each industry-year and then averaged across years. Error bars show 95% confidence intervals based on bootstrap (250 resamples).

Table 2: Average Markups for the Three Largest Industries in Each Country

	Industry 1				Industry 2				Industry 3			
	CD	TR	CES-FA	FA	CD	TR	CES-FA	FA	CD	TR	CES-FA	FA
					(311, 381, 321)							
Chile	1.35	1.33	1.29	1.21	1.43	1.42	1.39	1.28	1.32	1.29	1.31	1.27
	(0.01)	(0.01)	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)	(0.05)	(0.02)	(0.03)	(0.04)	(0.05)
					(311, 322, 381)							
Colombia	1.29	1.24	1.25	1.29	1.39	1.34	1.39	1.36	1.40	1.35	1.33	1.22
	(0.00)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)	(0.04)	(0.08)
					(230, 265, 213)							
India	1.2	1.2	1.14	1.15	1.17	1.16	1.17	1.16	1.42	1.44	1.33	1.26
	(0.01)	(0.02)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)
					(321, 311, 322)							
Turkey	1.26	1.11	1.19	1.16	1.37	1.31	1.29	1.24	1.30	1.22	1.34	1.32
	(0.01)	(0.05)	(0.02)	(0.02)	(0.00)	(0.05)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)
					(33, 32, 31)							
US	1.61	1.45	1.31	1.36	1.42	1.37	1.34	1.26	1.30	1.20	1.30	1.27
	(0.03)	(0.07)	(0.10)	(0.07)	(0.02)	(0.04)	(0.04)	(0.05)	(0.01)	(0.02)	(0.03)	(0.03)

Notes: This table reports sales-weighted average markup estimates across four production function specifications: Cobb-Douglas (CD), Translog (TR), CES with Hicks-neutral and labor-augmenting productivity (CES-FA), and strongly homothetic production function with Hicks-neutral and labor-augmenting productivity (FA). The estimates are presented for the three largest industries in each country. The average markups are calculated for each industry-year and then averaged over the years. Standard errors are derived using bootstrap (250 resamples). Industry codes are shown in parentheses, with corresponding names provided in Supplementary Appendix A.

CES-FA, and FA. Across all countries, the FA specification consistently yields lower markups than the CD specification, with differences ranging from 6.3 pp to 13.4 pp—a meaningful gap when markups are interpreted as measures of market power. Figure OA-4 in Online Appendix H, which reports the standard errors of these differences, confirms that these differences are statistically significant.[38] Specifications that are more flexible than CD—namely TR and CES-FA—demonstrate how functional form flexibility and labor-augmenting productivity reduce these biases. These specifications yield estimates between CD and FA values, reducing the biases by 15.2% to 86.5% across countries, but not eliminating them entirely.

Table 2, which presents markup estimates for the three largest industries in each country, reinforces these findings. In nearly all industries, the CD specification estimates the highest markups, while the FA specification estimates the lowest, with TR and CES-FA estimates falling in between. These estimates highlight the role of both functional-form flexibility and unobserved heterogeneity in production functions when estimating markups. Moreover, the consistency of patterns observed across various countries and industries indicates that the results are not driven by specific time periods or country- and industry-specific factors.

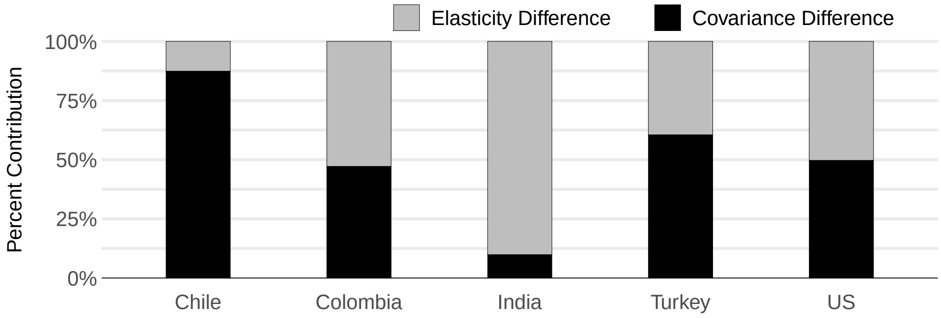
What mechanisms explain the differences in markup estimates across production functions? Section 7.2 highlighted two sources: (i) bias in average output elasticities and (ii) failure to capture the correlation between output elasticities and firm size. To quantify the relative importance of each mechanism, I partition the gap between CD and FA log-markup estimates ($\tilde{\mu_t}^{CD} - \tilde{\mu_t}^{FA}$) into these components using the decomposition method presented in Equation (7.2):

$$\Big(\mathbb{E}[\log(\theta_{it}^{CD})] - \mathbb{E}[\log(\theta_{it}^{FA})]\Big) + \Big(\text{Cov}\left(w_{it}, \log(\theta_{it}^{CD})\right) - \text{Cov}\left(w_{it}, \log(\theta_{it}^{FA})\right)\Big) \tag{8.1}$$

Figure 5 presents this decomposition, reporting the average elasticity differences (gray bars) and average covariance differences (black bars) as percentages of the total log-markup difference. Both components contribute roughly equally to markup differences on average (49.1% vs 50.9%), although their relative importance varies

[38]Figure OA-5 in Online Appendix H presents the year-by-year markup estimates, showing that the lower markup estimates from FA are not an artifact of averaging over time. The FA markup estimates are consistently below the CD estimates across all years throughout the sample.

Figure 5: Decomposition of the Difference between Aggregate Markups



Notes: This figure presents the decomposition results from Equation (8.1). The grey bars represent the percent contribution of elasticity differences (first term) to markup differences, while the black bars represent the percent contribution of covariance differences (second term). The decomposition is performed separately for each year and then averaged across years within each country.

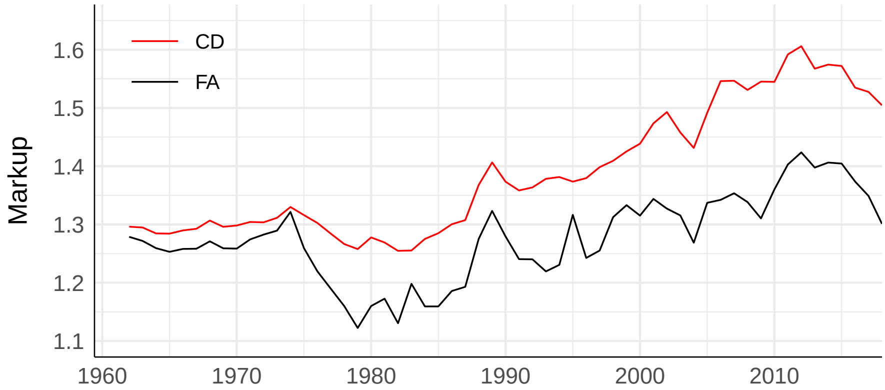
across countries. The results highlight two key patterns. First, the CD specification consistently overestimates the unweighted average flexible input elasticity across all countries (as indicated by the positive gray bars). Second, it fails to capture the negative relationship between firm size and flexible input elasticity (as indicated by the positive black bars). Markup bias therefore does not originate from a single source, emphasizing the importance of accurately estimating both the average elasticities in an industry and their heterogeneity across firms.

8.2 Markups Comparison: Trend

This section analyzes markup trends in US manufacturing, while corresponding estimates for other countries are reported in Figure OA-5 of Online Appendix H. The focus on the US is motivated by growing interest in using the production approach to analyze changes in market power (Autor et al. 2020; De Loecker et al. 2020). This line of research has generated substantial attention in both the economic implications of rising market power and methodologies for markup measurement (Berry et al. 2019; Miller 2025). I aim to contribute to this literature by exploring how a more flexible production function specification influences the estimates of markup changes.

Figure 6 plots the sales-weighted aggregate markup changes in US manufacturing from 1962 to 2018 based on CD and FA production functions. The FA estimates show that the aggregate markup remained relatively stable around 25% throughout the 1960s and mid-1970s, after which it experienced a gradual decline, falling to about 1.15 by 1980. Markups then began to increase again, exhibiting

Figure 6: Sales-Weighted Markup Trend Over Time in US Manufacturing



Notes: This figure compares sales-weighted markup trends in US manufacturing derived from a Cobb-Douglas (CD) production function versus a strongly homothetic production function with Hicks-neutral and factor-augmenting productivity (FA).

cyclical fluctuations and ultimately reaching around 1.3 by the end of the sample.

When we compare the markup estimates obtained from the FA and CD specifications, we observe that the CD estimates reveal an increase in markups of 26.4 pp (s.e.=3.1 pp) between the 1960s and 2010s. The FA specification also shows a rise in markups, though more modest at around 11.7 pp (s.e.=9.9 pp) over the same period.[39,40] The two series closely match up to the 1970s, after which they begin to diverge, with FA estimates peaking at 1.42 and CD estimates at 1.61 near the end of the sample period.

As described in Section 7, decomposing markup variation into components attributable to elasticities and revenue shares could provide insights into these differences. Figure OA-6 in Online Appendix H shows that variations in output elasticities account for only 1.3% of the variance in the aggregate markup under the CD specification, whereas they account for 21.5% under the FA specification. Thus, while revenue shares predominantly drive markup variation in both specifications, the FA approach indicates more variation in output elasticities over time than the CD specification.

These results complement recent findings on markups and market power in US

[39]The corresponding markup changes from the CES and TR specifications are 39.3 pp and 20.6 pp, respectively. The difference between the CD and FA markups is statistically significant at the 95% confidence level for most years after the 1980s, as reported in Figure OA-5 of Online Appendix H.

[40]The magnitude of markup changes is somewhat sensitive to the comparison period, though the direction remains consistent. Compared to the 1980s baseline period, when the FA estimate has the lowest value, CD estimates suggest a 25.2 pp (s.e.=2.3 pp) increase in the aggregate markup, whereas FA estimates suggest a 17.7 pp (s.e.=8.3 pp) increase.

manufacturing. For instance, Foster et al. (2024) find stable or declining markups using Census data when estimating the production function at the 4-digit industry level. Other studies have adopted demand-based methods to measure markups in various manufacturing industries. Grieco et al. (2023) find no rise in average markups in the auto industry, while others identify rising markups, including Miller et al. (2023) in the cement industry, Döpper et al. (2024) in consumer packaged goods, and De Loecker and Scott (2024) in the beer industry.

In concluding this section, it is important to note that these findings are specific to the manufacturing sector and should not be generalized to the entire economy. The drivers of market power likely vary across sectors, with industry-specific research showing distinct patterns of markups in different industries (Miller 2025). This paper's empirical results emphasize the importance of allowing for production function flexibility when measuring markups using the production approach. Moreover, this paper focuses on only one dimension of production function flexibility through labor-augmenting productivity. Incorporating other dimensions of flexibility—such as accounting for market power in input and output markets, multi-product production, or adjustment costs— could further improve our understanding of markup measurement (Doraszelski and Jaumandreu 2023; Cairncross et al. 2025; Rubens 2023; Ackerberg and De Loecker 2024).

9 Robustness Checks and Extensions

This section describes the robustness checks presented in Online Appendix G.

9.1 Quantity Production Functions

Most production datasets report revenue rather than physical output. In the presence of unobserved demand shocks and imperfect competition, revenue-based output elasticities may fail to accurately identify markups (Flynn et al. 2019; Bond et al. 2021). To analyze the robustness of my findings to this concern, I estimate production functions using physical output in six Indian manufacturing industries.

I focus on industries producing relatively homogeneous products: brick tiles, cotton shirts, biri cigarettes, black tea, parboiled non-basmati rice, and raw non-basmati rice.[41] Following Raval (2023), the sample includes plants that obtain at

[41] Using physical quantities in production functions presents its own challenges, such as account-

least 75% of their revenue from one of these products. Using this sample, I measure output in physical units and then repeat the estimation procedure described in Section 5. The results presented in Figure OA-7a of Online Appendix H indicate that relying on revenues instead of physical quantities introduces an upward bias in average capital elasticity (7.6%) and a downward bias in average labor elasticity (12.0%). However, the main findings in Section 8.1 on markup comparisons across production function models remain robust when using quantity data.

9.2 Heterogeneous Input Prices

Although my baseline empirical analysis assumes uniform input prices due to limitations of traditional production datasets, new datasets increasingly provide input price information. Not accounting for heterogeneous input prices may introduce biases into production function estimates. To address this concern, I utilize wage and intermediate input price data from a subset of Indian industries that predominantly rely on a single intermediate input. Following the extension outlined in Online Appendix A, I estimate output elasticities while controlling for input prices in the input demand functions. Results in Figure OA-7b of Online Appendix H show that in this empirical setting, failing to control for input prices results in small negative biases (0.3%–2.3%) in output elasticities and markups; however, the comparison findings across production function specifications remain robust. This analysis also serves as an example of how additional control variables can be incorporated into the estimation, which could be useful in other empirical contexts.

9.3 Multi-Product Firms

Another potential concern is that many firms across different industries produce multiple products, even though many production function estimation methods, including this paper, model single-product firms (Orr 2022; Valmari 2023). To evaluate the impact of this concern, I repeat the estimation procedure using only single-product firms from Indian manufacturing industries. Results in Figure OA-7c of Online Appendix H indicate some biases in output elasticities ranging from 76.4% for capital elasticity to -10.7% for labor elasticity. However, the markup

ing for quality differences, aggregating different units for multi-product firms, and comparing different units across firms. By focusing on homogeneous products, I aim to address these issues.

comparison findings from single-product firms are qualitatively similar to the main findings of the paper.

9.4 Measurement Error and Utilization in Capital

Another challenge in production function estimation is the potential measurement error in capital input and the unobservability of capital utilization, which can be exacerbated in nonparametric production functions (Collard-Wexler and De Loecker 2020). To assess whether these concerns explain this paper's findings, I perform two analyses described in Online Appendix G.3. First, a simulation exercise suggests that measurement error introduces a larger downward bias in capital elasticity for the labor-augmenting production function than for the Cobb-Douglas specification—a result inconsistent with my findings. Second, by assuming capital and electricity inputs are perfect complements, I derive capital utilization in Chilean and Turkish data and re-estimate the production function using utilized capital. While this correction affects the levels of elasticities and markups, the results are qualitatively similar to the main empirical findings.

10 Concluding Remarks

Production function estimation is central to analyzing various economic questions related to firm behavior. For an accurate understanding of firm behavior, it is important that our production functions capture the heterogeneity in production technologies among firms. This paper takes a step in this direction by developing a method for estimating production functions with factor-augmenting productivity and by demonstrating its impact on output elasticities and markup estimates.

Methodologically, I propose an approach that identifies output elasticities from a production function that includes both labor-augmenting and Hicks-neutral productivity. This approach imposes a functional form structure known as homothetic separability and exploits the cost minimization assumption to develop control variables for productivity shocks. Empirically, I demonstrate across five datasets that neglecting labor-augmenting productivity and imposing parametric restrictions can lead to biased estimates of output elasticities and markups.

A Proofs of Main Results

Proof of Proposition 2.1

This proof builds on the classic result of Shephard (1953). Throughout, I assume that the standard properties of production functions hold (Chambers (1988, p.9)), which guarantees that the cost function exists and Shephard's Lemma holds.

Part (i)

The firm minimizes the cost of flexible inputs to produce the planned output, $\bar{Y}_{it}$:

$$\min_{L_{it},M_{it}} p_t^l L_{it} + p_t^m M_{it} \quad \text{s.t.} \quad \mathbb{E}\big[F_t\left(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\right)\exp(\omega_{it}^H)\exp(\epsilon_{it}) \mid \mathcal{I}_{it}\big] \geqslant \bar{Y}_{it}.$$

Since the information set includes $(K_{it}, \omega_{it}^L, \omega_{it}^H)$, we can write this problem as:

$$\min_{L_{it},M_{it}} p_t^l L_{it} + p_t^m M_{it} \quad \text{s.t.} \quad F_t\left(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\right)\exp(\omega_{it}^H)\mathcal{E}_{it}(\mathcal{I}_{it}) \geqslant \bar{Y}_{it}, \tag{A.1}$$

where $\mathcal{E}_{it}(\mathcal{I}_{it}) := \mathbb{E}[\exp(\epsilon_{it}) \mid \mathcal{I}_{it}]$. Because the firm treats ω_{it}^L as fixed, this problem can be recast as a cost minimization problem where the firm chooses effective labor while facing quality-adjusted input prices. To see this, let $\bar{L}_{it} := \omega_{it}^L L_{it}$ denote the effective (quality-adjusted) labor and $\bar{p}_{it}^l := p_t^l/\omega_{it}^L$ denote the quality-adjusted wage. The cost minimization problem in Equation (A.1) can be rewritten as

$$\min_{M_{it},\bar{L}_{it}} \bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} \quad \text{s.t.} \quad F_t\big(K_{it}, h_t(K_{it}, \bar{L}_{it}, M_{it})\big)\exp(\omega_{it}^H) \geqslant \tilde{Y}_{it}(\mathcal{I}_{it}), \tag{A.2}$$

where $\tilde{Y}_{it}(\mathcal{I}_{it}) := \bar{Y}_{it}/\mathcal{E}_{it}(\mathcal{I}_{it})$. I will suppress $(\mathcal{I}_{it})$ in $\tilde{Y}_{it}$ since $\mathcal{E}_{it}(\mathcal{I}_{it})$ is a constant due to independence of ϵ_{it} from $\mathcal{I}_{it}$. Next, I derive the cost function from Equation (A.2). Letting $\bar{p}_{it} = (\bar{p}_{it}^l, p_t^m)$ denote the input prices, the cost function becomes:

$$\begin{aligned}
C_t(\tilde{Y}_{it}, K_{it}, &\omega_{it}^H, \bar{p}_{it}) \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : \tilde{Y}_{it} \leqslant F_t\big(K_{it}, h_t(K_{it}, \bar{L}_{it}, M_{it})\big)\exp(\omega_{it}^H)\Big\}, \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it} : F_t^{-1}(K_{it}, \tilde{Y}_{it}/\exp(\omega_{it}^H)) \leqslant h_t(K_{it}, \bar{L}_{it}, M_{it})\Big\}, \\
&= \min_{\bar{L}_{it},M_{it}} \Big\{F_t^{-1}(K_{it}, \tilde{Y}_{it}/\exp(\omega_{it}^H))\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h_t\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
&= F_t^{-1}(K_{it}, \tilde{Y}_{it}/\exp(\omega_{it}^H)) \min_{\bar{L}_{it},M_{it}} \Big\{\left(\bar{p}_{it}^l \bar{L}_{it} + p_t^m M_{it}\right) : 1 \leqslant h_t\big(K_{it}, \bar{L}_{it}, M_{it}\big)\Big\}, \\
&\equiv C_{1t}(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_{2t}(K_{it}, \bar{p}_{it}^l, p_t^m).
\end{aligned} \tag{A.3}$$

where $F_t^{-1}(\cdot)$ is the inverse function of $F(\cdot)$ with respect to its second argument. The second line follows from the assumption that $F_t(\cdot,\cdot)$ is strictly monotone in its second argument. In the third line, we exploit the homogeneity of

h_t and the linearity of the cost function to rewrite $h_t\big(K_{it}, F_t(\cdot)\bar{L}_{it}, F_t(\cdot)M_{it}\big)$ as $F_t(\cdot)h_t\big(K_{it}, \bar{L}_{it}, M_{it}\big)$. The fourth line follows because firms take $(K_{it}, \tilde{Y}_{it}, \omega^H_{it})$ as given. In the last line, the factorization in the previous step allows me to introduce two new functions—$C_{1t}(\cdot)$ and $C_{2t}(\cdot)$—which together describe the cost function.

By Shephard's Lemma, the firm's optimal demands for flexible inputs are given by the derivatives of the cost function with respect to the input prices. Applying this result yields the following ratio of materials to effective labor:

$$\frac{M_{it}}{\bar{L}_{it}} = \frac{\partial C_{2t}(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial p^m_t}{\partial C_{2t}(K_{it}, \bar{p}^l_{it}, p^m_t)/\partial \bar{p}^l_{it}} \equiv \frac{C_{mt}(K_{it}, \bar{p}^l_{it}, p^m_t)}{C_{lt}(K_{it}, \bar{p}^l_{it}, p^m_t)},$$

where $C_{mt}(\cdot)$ and $C_{lt}(\cdot)$ are the derivatives of the function $C_{2t}(\cdot)$ with respect to p^m_t and $\bar{p}^l_{it}$, respectively. Because $\bar{L}_{it} = \omega^L_{it}L_{it}$, we can express M_{it}/L_{it} as:

$$\frac{M_{it}}{L_{it}} = \frac{C_{mt}(K_{it}, \bar{p}^l_{it}, p^m_t)\omega^L_{it}}{C_{lt}(K_{it}, \bar{p}^l_{it}, p^m_t)}. \tag{A.4}$$

The resulting function depends on K_{it}, ω^L_{it}, and industry-specific input prices:

$$\tilde{M}_{it} = \bar{r}_t(K_{it}, \omega^L_{it}, p^m_t, p^l_t) \equiv r_t(K_{it}, \omega^L_{it}), \tag{A.5}$$

for some function $r_t(K_{it}, \omega^L_{it})$. This completes the first part of the proof.

Part (ii)

In the second part of the proof, I will show that

$$\partial r_t(K_{it}, \omega^L_{it})/\partial \omega^L_{it} > 0 \quad \text{for all } (K_{it}, \omega^L_{it}) \quad \text{or} \quad \partial r_t(K_{it}, \omega^L_{it})/\partial \omega^L_{it} < 0 \quad \text{for all } (K_{it}, \omega^L_{it}).$$

By the properties of the cost functions, $C_{mt}(\cdot)$ and $C_{lt}(\cdot)$ are homogeneous of degree zero in input prices (Chambers (1988, p.64)). Thus, we can divide all input prices by p^m_t in Equation (A.4) and express the input ratio as:

$$\tilde{M}_{it} = \frac{C_{mt}(K_{it}, \bar{p}^l_{it}/p^m_t, 1)\omega^L_{it}}{C_{lt}(K_{it}, \bar{p}^l_{it}/p^m_t, 1)} \equiv \frac{\tilde{C}_{mt}(K_{it}, p^{l/m}_{it})\omega^L_{it}}{\tilde{C}_{lt}(K_{it}, p^{l/m}_{it})}, \tag{A.6}$$

where $p^{l/m}_{it} := \bar{p}^l_{it}/p^m_t$, $\tilde{C}_{mt}(K_{it}, p^{l/m}_{it}) := C_{mt}(K_{it}, p^{l/m}_{it}, 1)$, and $\tilde{C}_{lt}(K_{it}, p^{l/m}_{it}) := C_{lt}(K_{it}, p^{l/m}_{it}, 1)$. Taking the logarithm of Equation (A.6) yields

$$\log(\tilde{M}_{it}) = -\log\Big(\tilde{C}_{lt}(K_{it}, p^{l/m}_{it})/\tilde{C}_{mt}(K_{it}, p^{l/m}_{it})\Big) + \log(\omega^L_{it}).$$

Differentiating $\log(\tilde{M}_{it})$ with respect to $\log(\omega^L_{it})$ gives

$$\begin{aligned}
\frac{\partial \log(\tilde{M}_{it})}{\partial \log(\omega^L_{it})} &= -\frac{\partial \log\big(\tilde{C}_{lt}(K_{it}, p^{l/m}_{it})/\tilde{C}_{mt}(K_{it}, p^{l/m}_{it})\big)}{\partial \log(p^{l/m}_{it})}\left(\frac{\partial \log(p^{l/m}_{it})}{\partial \log(\omega^L_{it})}\right) + 1, \\
&= \frac{\partial \log\big(\tilde{C}_{lt}(K_{it}, p^{l/m}_{it})/\tilde{C}_{mt}(K_{it}, p^{l/m}_{it})\big)}{\partial \log(p^{l/m}_{it})} + 1 \equiv -\sigma_t(K_{it}, \omega^L_{it}L_{it}, M_{it}) + 1. \quad \text{(A.7)}
\end{aligned}$$

The last equality uses the fact that the elasticity of substitution equals minus the derivative of the log input ratio with respect to the log input-price ratio (Chambers (1988, p.94)).[42] By Assumption 2.1(iv), $\sigma_t(\cdot) < 1$ everywhere or $\sigma_t(\cdot) > 1$ everywhere. Thus, the flexible input ratio is strictly monotone in ω_{it}^L. $\square$

Lemma A.1. *Let x, y, and z be scalar continuous random variables with joint density $f(x,y,z)$. Assume (x,y) are jointly independent of z. Then x and z are independent conditional on y.*

Proof. Let $f(x \mid y)$ denote the conditional probability density function of x given y. The independence assumption implies that $f(x,y,z) = f(x,y)f(z)$. To complete the proof, it suffices to show that $f(x,z \mid y) = f(x \mid y)f(z \mid y)$. Using Bayes' rule for continuous random variables, I obtain

$$
\begin{aligned}
f(x,z \mid y) &= \frac{f(x,y,z)}{f(y)} = \frac{f(x,y)f(z)}{f(y)} = \frac{f(x \mid y)f(y)f(z)}{f(y)} = f(x \mid y)f(z), \\
&= f(x \mid y)f(z \mid y),
\end{aligned}
$$

where $f(z \mid y) = f(z)$ because joint independence of (x,y) from z implies that y is independent of z. $\square$

Proof of Lemma 3.1

By Assumption 2.3, we have that $\omega_{it}^L \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H$. Substituting ω_{it}^L from Equation (3.1), I obtain

$$
g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{A.8}
$$

Since $g(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$ is strictly monotone in u_{it}^1, Equation (A.8) implies

$$
u_{it}^1 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H. \tag{A.9}
$$

By normalization, u_{it}^1 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H)$ and by timing assumption $(K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H) \in \mathcal{I}_{it-1}$. Thus, Equation (A.9) implies

$$
u_{it}^1 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H \sim \mathrm{U}(0,1).
$$

$(\omega_{it-1}^L, \omega_{it-1}^H)$ are functions of W_{it-1} by Equations (2.6) and (2.7). Using this

$$
u_{it}^1 \mid K_{it}, W_{it-1}, \bar{r}_{t-1}(K_{it-1}, \tilde{M}_{it-1}), \bar{s}_{t-1}(K_{it-1}, M_{it-1}, \tilde{M}_{it-1}) \sim \mathrm{U}(0,1),
$$

[42] To see this, note that the elasticity of substitution between inputs (X_1, X_2) with prices (p_1, p_2) is defined as $\sigma = \partial \log(X_1/X_2)/\partial \log(F_1/F_2)$. By cost minimization, $F_1/F_2 = p_1/p_2$, thus $\sigma = -\partial \log(X_1/X_2)/\partial \log(p_1/p_2)$.

which implies $u_{it}^1 \mid K_{it}, W_{it-1} \sim \mathrm{U}(0,1)$. $\square$

Proof of Lemma 3.2

By Assumption 2.3, we have that $(\omega_{it}^L, \omega_{it}^H) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H$. Using the representations of productivity shocks in Equations (3.1) and (3.5) yields

$$g_L(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1), g_H(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1, u_{it}^2) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H.$$

Monotonicity of $g_L(\cdot)$ in u_{it}^1, monotonicity $g_H(\cdot)$ in u_{it}^2, Lemma A.1 imply that

$$u_{it}^2 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1. \tag{A.10}$$

It follows from Equation (A.10), the fact that u_{it}^2 is uniformly distributed conditional on $(\omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1)$, and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ that

$$u_{it}^2 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, u_{it}^1 \sim \mathrm{U}(0,1).$$

$(\omega_{it-1}^L, \omega_{it-1}^H)$ are functions of W_{it-1} by Equations (2.6)-(2.7). Using this, we obtain

$$u_{it}^2 \mid K_{it}, W_{it-1}, u_{it}^1 \sim \mathrm{U}(0,1). \qquad \square$$

B Identification of Reduced Form Production Function

In this section, I show that the reduced form representation of strong homothetic production $(f_t, \bar{h}_t)$ in Equation (4.7) is identified using the moment in Equation (5.3). I begin by proving auxiliary lemmas and then present the main proof.

B.1 Auxiliary Lemmas

Lemma B.1. *Suppose that w, z, x, s are continuous random variables and that* $\mathrm{Var}\big(s \mid w, z, x\big) > 0$. *Then, the functions $f(w,z,x)$ and $h(w,z,x)$ can be uniquely identified from the relationship*

$$y(w,z,x,s) = f(w,z,x) - h(w,z,x)s.$$

where $y(w,z,x,s)$ is a known function.

Proof. Given that $\mathrm{Var}\big(s \mid w, z, x\big) > 0$, it follows that there exist at least two distinct values of s, say s_1 and s_2, with positive probability for each fixed (w,z,x). Considering these two distinct values of s for the same (w,z,x), we can write

$$y_1 = f(w,z,x) - h(w,z,x)s_1, \tag{B.1}$$

$$y_2 = f(w,z,x) - h(w,z,x)s_2. \tag{B.2}$$

Subtract Equation (B.1) from Equation (B.2) to eliminate $f(w, z, x)$:

$$y_2 - y_1 = h(w, z, x)(s_1 - s_2).$$

which identifies $h(w, z, x)$ as: $h(w, z, x) = (y_2 - y_1)/(s_1 - s_2)$. Substituting $h(\cdot)$ into Equation (B.1) and solving for $f(w, z, x)$ identifies $f(w, z, x)$. □

Lemma B.2. *Let $p : \mathbb{R}_+^3 \to \mathbb{R}$ be a known, differentiable function. Suppose there exist unknown differentiable functions $f : \mathbb{R}_+^2 \to \mathbb{R}$ and $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $w, x, z > 0$, $f(w, zh(x)) = p(w, x, z)$, then $h(x)$ is identified up to a multiplicative constant, and $f(w, y) := f(w, zh(x))$ (where $y := zh(x)$) is identified up to a normalization given in the proof.*

Proof. Taking derivatives of $f(w, zh(x)) = p(w, x, z)$ with respect to z and x yields

$$f_2(w, zh(x))h(x) = p_3(w, x, z), \qquad f_2(w, zh(x))zh'(x) = p_2(w, x, z),$$

where the subscript k denotes the derivative of the function with respect to the k^{th} argument, e.g., $f_2(w, zh(x)) = \partial f(w, y)/\partial y$. Taking the ratio of these derivatives:

$$\frac{h'(x)}{h(x)} = \frac{p_2(w, x, z)}{p_3(w, x, z)z}. \tag{B.3}$$

Recognizing that $h'(x)/h(x) = (d/dx)\log h(x)$, we can identify $\log(h(x))$ up to a constant by integrating Equation (B.3). Hence, $h(x)$ is identified up to a scale. Then, $f(w, y)$ is identified from $f(w, zh(x)) = p(w, x, z)$ up to the following normalization: $\tilde{f}(w, y) = f(w, y/c)$ and $\tilde{h}(x) = ch(x)$ with $c > 0$ such that if (f, h) satisfy the conditions of the proposition then $(\tilde{f}, \tilde{h})$ satisfy them too. □

Lemma B.3. *Suppose $f(x, y)$ has continuous partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$. Then it is uniquely determined up to a constant by its derivatives.*

Proof. Define $f_1(\cdot) = \partial f(x, y)/\partial x$ and $f_2(\cdot) = \partial f(x, y)/\partial y$. Integrating $f_1(x, y)$:

$$f(x, y) = \int f_1(x, y)\, dx + g(y), \tag{B.4}$$

where $g(y)$ is an arbitrary function. Differentiating Equation (B.4) w.r.t. y:

$$f_2(x, y) = \frac{\partial}{\partial y}\left(\int f_1(x, y)\, dx\right) + g_1(y).$$

Solving for $g_1(y)$:

$$g_1(y) = f_2(x, y) - \frac{\partial}{\partial y}\left(\int f_1(x, y)\, dx\right). \tag{B.5}$$

Integrate Equation (B.5) with respect to y:

$$g(y) = \int \left[f_2(x,y) - \frac{\partial}{\partial y}\left(\int f_1(x,y)\,dx \right) \right] dy + C, \tag{B.6}$$

where C is a constant. Substituting Equation (B.6) into Equation (B.4):

$$f(x,y) = \int f_1(x,y)\,dx + \int \left[f_2(x,y) - \frac{\partial}{\partial y}\left(\int f_1(x,y)\,dx \right) \right] dy + C. \tag{B.7}$$

Thus, $f(x,y)$ is determined up to a constant by its partial derivatives. □

B.2 Proof of Proposition 5.1

The strongly homothetic production function has the following reduced-form:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it}. \tag{B.8}$$

Substituting the control function for ω_{it}^H we obtain:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it} \mid K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1}, u_{it}^1, u_{it}^2] = 0.$$

Under strongly homothetic production function, the control variables are $u_{it}^1 = F_{\tilde{M}_{it}}(\tilde{M}_{it})$ and $u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. Since $F_{\tilde{M}_{it}}(\tilde{M}_{it})$ is a monotone transformation of $\tilde{M}_{it}$, we can set u_{it}^1 to $\tilde{M}_{it}$ for the purposes of this proof. Substituting u_{it}^1 and u_{it}^2 into the equation above gives:

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}\big(W_{it-1}, \tilde{M}_{it}, \tilde{s}_t(K_{it}, M_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it},$$

where $\tilde{s}_t(\cdot) := F_{M_{it}|K_{it},W_{it-1},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, u_{it}^1)$. In this equation, $(f_t, \bar{h}_t, c_{2t})$ are unknown functions to be estimated, and $\tilde{s}_t(\cdot)$ can be treated as known in this proof since it is identified from the data in the first stage.

I will next simplify the notation to present the proof in a compact form. First I transform the arguments of $\tilde{s}_t(\cdot)$ and can rewrite this equation as

$$y_{it} = f_t\big(K_{it}, L_{it}\bar{h}_t(\tilde{M}_{it})\big) + c_{2t}\big(W_{it-1}, \tilde{M}_{it}, s_t(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})\big) + \epsilon_{it}, \tag{B.9}$$

where $\tilde{s}_t(x_1, x_2, x_3, x_4) = s_t(x_1, x_2/x_3, x_3, x_4)$. Second, I relabel $(K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})$ as (w, z, x, t), $\bar{h}_t$ as h, c_{2t} as g, and drop the time and firm subscripts to obtain

$$y = f\big(w, zh(x)\big) + g\big(t, x, s(w,z,x,t)\big) + \epsilon, \qquad \mathbb{E}[\epsilon \mid w, z, x, t] = 0.$$

Taking expectation of both sides conditional on (w, z, x, t), we obtain

$$\mathbb{E}[y \mid w, z, x, t] = f\big(w, zh(x)\big) + g\big(t, x, s(w,z,x,t)\big). \tag{B.10}$$

Since the data identifies $\mathbb{E}[y \mid w, z, x, t]$ as a conditional expectation we must show that f and h are identified from $\mathbb{E}[y \mid w, z, x, t]$. Let Ω be the space of functions satisfying our assumptions, where true functions $(f_0, h_0, g_0) \in \Omega$. Following Matzkin (2007), $(\tilde{f}, \tilde{h}, \tilde{g}) \in \Omega$ is *observationally equivalent* to (f_0, h_0, g_0) if

$$f_0\big(w, zh_0(x)\big) + g_0\big(t, x, s(w, z, x, t)\big) = \tilde{f}\big(w, z\tilde{h}(x)\big) + \tilde{g}\big(t, x, s(w, z, x, t)\big). \quad \text{(B.11)}$$

$(f_0, h_0, g_0) \in \Omega$ are identifiable if no other member of Ω is observationally equivalent to them. If identification holds except in special or pathological cases, then the model is said to be *generically identified* (Lewbel 2019).

Proposition B.1. *Suppose that (i) Functions* (f_0, h_0, g_0, s) *are twice continuously differentiable and have non-zero derivatives almost everywhere, (ii) The joint distribution function of* (w, z, x, t) *is absolutely continuous with positive density everywhere on its support, (iii)* $\mathrm{Var}(s_1(w, z, x, t)/s_2(w, z, x, t) \mid w, z, x) > 0$ *where* $s_1 \equiv \partial s/\partial w$ *and* $s_2 \equiv \partial s/\partial z$*. Then* g_0 *is identified up to a constant,* h_0 *is identified up to a scale, and* f_0 *is identified up to a normalization given in the proof.*

Proof. I will show that if there exists observationally equivalent (f, h, g) and $(\tilde{f}, \tilde{h}, \tilde{g})$ as defined in Equation (B.11), then they must equal each other up to the normalization given in the proposition. Let $y(w, z, x, t)$ denote $\mathbb{E}[y \mid w, x, z, t]$. Taking derivatives of $y(w, z, x, t)$ with respect to w and z using Equation (B.10) yields the following equations:

$$y_1(w, z, x, t) = f_1\big(w, zh(x)\big) + g_3\big(t, x, s(w, z, x, t)\big)s_1(w, z, x, t), \quad \text{(B.12)}$$

$$y_2(w, z, x, t) = f_2\big(w, zh(x)\big)h(x) + g_3\big(t, x, s(w, z, x, t)\big)s_2(w, z, x, t), \quad \text{(B.13)}$$

where a function subscript k (e.g., f_1) denotes the partial derivative of that function with respect to its k-th argument. Multiplying Equation (B.13) by $s_1(w, z, x, t)/s_2(w, z, x, t)$ and subtracting equation (B.12) yields

$$\begin{aligned} y_1(w,z, x, t) - y_2(w, z, x, t)\big(s_1(w, z, x, t)/s_2(w, z, x, t)\big) = \\ f_1\big(w, zh(x)\big) - f_2\big(w, zh(x)\big)h(x)\big(s_1(w, z, x, t)/s_2(w, z, x, t)\big) \end{aligned} \quad \text{(B.14)}$$

The left-hand side of Equation (B.14) includes identified functions. Denote it by $\widehat{y}(w, z, x, t)$ and define $\widehat{f_1}(w, z, x) := f_1\big(w, zh(x)\big)$ and $\widehat{f_2}(w, z, x) := f_2\big(w, zh(x)\big)h(x)$. With these functions, Equation (B.14) can be written as:

$$\widehat{y}(w, z, x, t) = \widehat{f_1}(w, z, x) - \widehat{f_2}(w, z, x)\widehat{s}(w, z, x, t) \quad \text{(B.15)}$$

where $\widehat{s}(\cdot) = s_1(\cdot)/s_2(\cdot)$ By Assumption (iii), applying Lemma B.1 implies that both $\widehat{f}_1(w, z, x)$ and $\widehat{f}_2(w, z, x)$ are identified from Equation (B.15). To separately identify $h(x)$ and $f_1\big(w, v\big) := f_1\big(w, zh(x)\big)$ from $\widehat{f}_1(w, z, x)$, we apply Lemma B.2. This lemma shows that $h(x)$ is identified up to scale ($h(x) = c\tilde{h}(x)$) and $f_1\big(w, v\big)$ is identified from $\widehat{f}_1(w, z, x)$ up to a normalization ($\tilde{f}(w, v) = f(w, v/c)$). Once we know $h(x)$, the identification of $\widehat{f}_2(w, z, x)$ implies that $f_2\big(w, v\big)$ is also identified up to the same normalization.

To proceed with the results, assume that c is fixed. Applying Lemma B.3 shows that the identification of its partial derivatives implies the identification of $f(w, v)$ up to a constant. Therefore, for a fixed c, $h(x)$ is identified, and $f(w, v)$ is identified up to a constant. Since $c > 0$ can take any positive value, we have that

$$f(w, v) = \tilde{f}(w, v/c) + a, \qquad h(x) = c\tilde{h}(x).$$

Here, c and a are arbitrary constants representing the scale and additive shifts. Therefore, all observationally equivalent functions satisfy these relationships, which concludes the proof of Proposition B.1. $\square$

To finish the proof of Proposition 5.1, we need to map its conditions to the conditions of Proposition B.1 as they are given in a different notation. The equivalence of Conditions (i-ii) under Proposition 5.1 and Proposition B.1 is trivial. For Condition (iii), Proposition 5.1 stated that $\mathrm{Var}(\tilde{s}_1(x_1, x_2, x_3, x_4)/\tilde{s}_2(x_1, x_2, x_3, x_4) \mid x_1, x_2, x_3) > 0$ where $\tilde{s}(x_1, x_2, x_3, x_4) = s(x_1, x_2/x_3, x_3, x_4)$. This implies Proposition B.1 Condition (iii) since:

$$\frac{\tilde{s}_1(x_1, x_2, x_3, x_4)}{\tilde{s}_2(x_1, x_2, x_3, x_4)} x_3 = \frac{s_1(x_1, x_2/x_3, x_3, x_4)}{s_2(x_1, x_2/x_3, x_3, x_4)}$$

based on the notation redefinition given in Equation (B.9). Taking the conditional variance implies that condition (iii) of Proposition 5.1 is equivalent to condition (iii) of Proposition B.1. Therefore, Proposition B.1 implies that Proposition 5.1 holds.

References

Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451.

Ackerberg, Daniel A. and De Loecker, Jan. 2024. "Production Function Identification Under Imperfect Competition". *Working Paper.*

Ackerberg, Daniel A., Hahn, Jinyong, and Pan, Qingsong. 2023. "Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity Shocks". *Working Paper.*

Anderson, Simon P., Erkal, Nisvan, and Piccinin, Daniel. 2020. "Aggregative Games and Oligopoly Theory: Short-Run and Long-Run Analysis". *The RAND Journal of Economics* 51 (2): 470–495.

Autor, David, Dorn, David, Katz, Lawrence F, Patterson, Christina, and Van Reenen, John. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms". *The Quarterly Journal of Economics* 135 (2): 645–709.

Balat, Jorge, Brambilla, Irene, and Sasaki, Yuya. 2022. "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports". *Working Paper.*

Basu, Susanto. 2019. "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence". *Journal of Economic Perspectives* 33 (3): 3–22.

Berry, Steven, Gaynor, Martin, and Scott Morton, Fiona. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization". *Journal of Economic Perspectives* 33 (3): 44–68.

Berry, Steven, Levinsohn, James, and Pakes, Ariel. 1995. "Automobile Prices in Market Equilibrium". *Econometrica*: 841–890.

Biondi, Filippo. 2022. "Firm Productivity and Derived Factor Demand Under Variable Markups". *Working Paper.*

Blackwood, G Jacob, Foster, Lucia S, Grim, Cheryl A, Haltiwanger, John, and Wolf, Zoltan. 2021. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights". *American Economic Journal: Macroeconomics* 13 (3): 142–172.

Bloom, Nicholas and Van Reenen, John. 2010. "Why Do Management Practices Differ Across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–224.

Blundell, Richard and Bond, Stephen. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340.

Bond, Steve, Hashemi, Arshia, Kaplan, Greg, and Zoch, Piotr. 2021. "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data". *Journal of Monetary Economics* 121:1–14.

Cairncross, John, Morrow, Peter, Orr, Scott, and Rachapalli, Swapnika. 2025. "Multi-product Markups". *SSRN Working Paper 4903621.*

Caplin, Andrew and Nalebuff, Barry. 1991. "Aggregation and Imperfect Competition: On the Existence of Equilibrium". *Econometrica* 59 (1): 25–59.

Caradonna, Peter, Miller, Nathan, and Sheu, Gloria. 2025. "Mergers, Entry, and Consumer Welfare". Forthcoming, *American Economic Review: Microeconomics.*

Chambers, Robert G. 1988. *Applied Production Analysis: A Dual Approach.* Cambridge University Press.

Chernozhukov, Victor and Hansen, Christian. 2008. "Instrumental Variable Quantile Regression: A Robust Inference Approach". *Journal of Econometrics* 142 (1): 379–398.

Clemhout, Simone. 1968. "The Class of Homothetic Isoquant Production Functions". *The Review of Economic Studies* 35 (1): 91–104.

Collard-Wexler, Allan and De Loecker, Jan. 2020. "Production Function Estimation and Capital Measurement". *NBER Working Paper w22437.*

De Loecker, Jan. 2011. "Recovering Markups from Production Data". *International Journal of Industrial Organization* 29 (3): 350–355.

De Loecker, Jan, Eeckhout, Jan, and Unger, Gabriel. 2020. "The Rise of Market Power and the Macroeconomic Implications". *The Quarterly Journal of Economics* 135 (2): 561–644.

De Loecker, Jan and Scott, Paul. 2024. "Markup Estimation using Production and Demand Data. An Application to the US Brewing Industry". *Working Paper.*

De Loecker, Jan and Syverson, Chad. 2021. "An Industrial Organization Perspective on Productivity". In *Handbook of Industrial Organization*, 4:141–223. Elsevier.

De Loecker, Jan and Warzynski, Frederic. 2012. "Markups and Firm-Level Export Status". *American Economic Review* 102 (6): 2437–2471.

De Ridder, Maarten, Grassi, Basile, and Morzenti, Giovanni. 2025. "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data". *LSE Working Paper.*

Diamond, Peter, McFadden, Daniel, and Rodriguez, Miguel. 1978. "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change", 2:125–147.

Döpper, Hendrik, MacKay, Alexander, Miller, Nathan H, and Stiebale, Joel. 2024. "Rising Markups and the Role of Consumer Preferences". *NBER Working Paper w32739.*

Doraszelski, Ulrich and Jaumandreu, Jordi. 2013. "R&D and Productivity: Estimating Endogenous Productivity". *Review of Economic Studies* 80 (4): 1338–1383.

— . 2018. “Measuring the Bias of Technological Change”. *Journal of Political Economy* 126 (3): 1027–1084.

— . 2019. “Using Cost Minimization to Estimate Markups”. *Working Paper.*

— . 2023. “Reexamining the De Loecker and Warzynski (2012) Method for Estimating Markups”. *Working Paper.*

Dunne, Timothy, Haltiwanger, John, and Troske, Kenneth R. 1997. “Technology and Jobs: Secular Changes and Cyclical Dynamics”. In *Carnegie-Rochester Conference Series on Public Policy*, 46:107–178. Elsevier.

Flynn, Zach, Gandhi, Amit, and Traina, James. 2019. “Measuring Markups with Production Data”. *SSRN Working Paper 3358472.*

Foster, Lucia, Haltiwanger, John, and Tuttle, Cody. 2024. “Rising Markups or Changing Technology”. *NBER Working Paper w30491.*

Gandhi, Amit, Navarro, Salvador, and Rivers, David A. 2020. “On the Identification of Gross Output Production Functions”. *Journal of Political Economy* 128 (8): 2973–3016.

Garrido, Francisco. 2022. “An Aggregative Approach to Pricing Equilibrium Among Multi-Product Firms with Nested Demand”. *The RAND Journal of Economics.*

Grieco, Paul L E, Murry, Charles, and Yurukoglu, Ali. 2023. “The Evolution of Market Power in the U.S. Automobile Industry”. *The Quarterly Journal of Economics* 139 (2): 1201–1253.

Grieco, Paul L. E., Li, Shengyu, and Zhang, Hongsong. 2016. “Production Function Estimation with Unobserved Input Price Dispersion”. *International Economic Review* 57 (2): 665–690.

Hall, Robert E. 1988. “The Relation between Price and Marginal Cost in U.S. Industry”. *Journal of Political Economy* 96 (5): 921–947.

Haltiwanger, John, Kulick, Robert B, and Syverson, Chad. 2018. “Misallocation Measures: The Distortion That Ate the Residual”. *Becker Friedman Institute for Economics Working Paper.*

Hsieh, Chang-Tai and Klenow, Peter J. 2009. “Misallocation and Manufacturing TFP in China and India”. *The Quarterly Journal of Economics* 124 (4): 1403–1448.

Imbens, Guido and Newey, Whitney. 2009. “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity”. *Econometrica* 77 (5): 1481–1512.

Jensen, Martin Kaae. 2018. “Aggregative Games”. In *Handbook of Game Theory and Industrial Organization, Volume I*, 66–92. Edward Elgar Publishing.

Jorgenson, Dale W. 1986. “Econometric Methods for Modeling Producer Behavior”. In *Handbook of Econometrics*, 3:1841–1915. Elsevier.

Kasahara, Hiroyuki, Schrimpf, Paul, and Suzuki, Michio. 2023. “Identification and Estimation of Production Function with Unobserved Heterogeneity”. *arXiv preprint, 2305.12067.*

Kasy, Maximilian. 2011. “Identification in Triangular Systems Using Control Functions”. *Econometric Theory* 27 (3): 663–671.

Kehrig, Matthias and Vincent, Nicolas. 2021. “The Micro-level Anatomy of the Labor Share Decline”. *The Quarterly Journal of Economics* 136 (2): 1031–1087.

Keller, Wolfgang and Yeaple, Stephen R. 2009. “Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States”. *The Review of Economics and Statistics* 91 (4): 821–831.

Kmenta, Jan. 1967. “On Estimation of the CES Production Function”. *International Economic Review* 8 (2): 180–189.

Levinsohn, James and Petrin, Amil. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables”. *The Review of Economic Studies* 70 (2): 317–341.

Lewbel, Arthur. 2019. “The Identification Zoo: Meanings of Identification in Econometrics”. *Journal of Economic Literature* 57 (4): 835–903.

Lewbel, Arthur and Linton, Oliver. 2007. “Nonparametric Matching and Efficient Estimators of Homothetically Separable Functions”. *Econometrica* 75 (4): 1209–1227.

Li, Tong and Sasaki, Yuya. 2017. “Constructive Identification of Heterogeneous Elasticities in the Cobb-Douglas Production Function”. *arXiv:1711.10031.*

Matzkin, Rosa L. 2007. “Nonparametric Identification”. In *Handbook of Econometrics*, 6:5307–5368. Elsevier.

Miller, Nathan H. 2025. “Industrial Organization and the Rise of Market Power”. *International Journal of Industrial Organization* 98:103131.

Miller, Nathan, Osborne, Matthew, Sheu, Gloria, and Sileo, Gretchen. 2023. “Technology and Market Power: The United States Cement Industry, 1974-2019”. *Georgetown McDonough School of Business Research Paper*, no. 4041168.

Morlacco, Monica. 2020. “Market Power in Input Markets: Theory and Evidence from French Manufacturing”. *Working Paper.*

Nadiri, M. Ishaq. 1982. “Producers Theory”. In *Handbook of Mathematical Economics*, 2:431–490. Elsevier.

Nocke, Volker and Schutz, Nicolas. 2018. “Multiproduct-firm Oligopoly: An Aggregative Games Approach”. *Econometrica* 86 (2): 523–557.

Nocke, Volker and Whinston, Michael D. 2022. “Concentration Thresholds for Horizontal Mergers”. *American Economic Review* 112 (6): 1915–1948.

Oberfield, Ezra and Raval, Devesh. 2021. “Micro Data and Macro Technology”. *Econometrica* 89 (2): 703–732.

Olley, G. Steven and Pakes, Ariel. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry”. *Econometrica* 64 (6): 1263–1297.

Orr, Scott. 2022. “Within-Firm Productivity Dispersion: Estimates and Implications”. *Journal of Political Economy* 130 (11): 2771–2828.

Pan, Qingsong. 2024. “Identification of Gross Output Production Functions with a Nonseparable Productivity Shock”. *SSRN Working Paper 4899485.*

Raval, Devesh. 2011. “Beyond Cobb-Douglas: Estimation of a CES Production Function with Factor Augmenting Technology”. *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-05.*

— . 2019. “The Micro Elasticity of Substitution and Non-neutral Technology”. *The RAND Journal of Economics* 50 (1): 147–167.

— . 2023. “Testing the Production Approach to Markup Estimation”. *Review of Economic Studies* 90 (5): 2592–2611.

Rubens, Michael. 2023. “Market Structure, Oligopsony Power, and Productivity”. *American Economic Review* 113 (9): 2382–2410.

Rubens, Michael, Wu, Yingjie, and Xu, Mingzhi. 2025. “Exploiting or Augmenting Labor?” *Working Paper.*

Shapiro, Carl and Yurukoglu, Ali. 2025. “Trends in Competition in the United States: What Does the Evidence Show?” *Working Paper.*

Shephard, Ronald W. 1953. *Cost and Production Functions.* Princeton University Press.

Syverson, Chad. 2011. “What Determines Productivity?” *Journal of Economic Literature* 49 (2): 326–365.

Valmari, Nelli. 2023. “Estimating Production Functions of Multiproduct Firms”. *Review of Economic Studies* 90 (6): 3315–3342.

Van Beveren, Ilke. 2012. “Total Factor Productivity Estimation: A Practical Review”. *Journal of Economic Surveys* 26 (1): 98–128.

Van Biesebroeck, Johannes. 2008. “The Sensitivity of Productivity Estimates: Revisiting Three Important Debates”. *Journal of Business & Economic Statistics* 26 (3): 311–328.

Van Reenen, John. 2018. “Increasing Differences between Firms: Market Power and the Macro-Economy”. *Federal Reserve Bank of Kansas City Working Paper.*

Wooldridge, Jeffrey M. 2009. “On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables”. *Economics Letters* 104 (3): 112–114.

Yeh, Chen, Macaluso, Claudia, and Hershbein, Brad. 2022. “Monopsony in the US Labor Market”. *American Economic Review* 112 (7): 2099–2138.

Zhang, Hongsong. 2019. “Non-neutral Technology, Firm Heterogeneity, and Labor Demand”. *Journal of Development Economics* 140 (3): 145–168.

Production Function Estimation with Factor-Augmenting Technology: An Application to Markups

Mert Demirer

Appendix - For Online Publication

A Extension to Heterogeneous Input Prices

In this extension, I allow input prices to vary across firms. Let p_{it}^{l} and p_{it}^{m} denote the wage and materials prices. Define $\bar{p}_{it} := (p_{it}^{l}, p_{it}^{m})$ as the input price vector and define $p_{it}^{l/m} := p_{it}^{l}/p_{it}^{m}$ as the input price ratio. Unlike in the main model, we now include input prices in W_{it}, so we let $W_{it} := (K_{it}, L_{it}, M_{it}, \bar{p}_{it})$. To account for firm-level variation in input prices, I revise the Markov and monotonicity assumptions as follows.

Assumption OA-1 (First-Order Markov with Heterogeneous Input Prices)**.** *The distribution of productivity shocks and input prices satisfies the following condition:*

$$P(\omega_{it}^{L}, \omega_{it}^{H}, \bar{p}_{it} \mid \mathcal{I}_{it-1}) = P(\omega_{it}^{L}, \omega_{it}^{H}, \bar{p}_{it} \mid \omega_{it-1}^{L}, \omega_{it-1}^{H}, \bar{p}_{it-1}).$$

This assumption states that input prices and productivity shocks jointly follow an exogenous, first-order Markov process. Notably, this assumption allows productivity shocks and input prices to be correlated, capturing cases such as more productive workers (those with higher ω_{it}^{L}) earning higher wages.

Assumption OA-2 (Monotonicity with Heterogeneous Input Prices)**.** *The firm's materials demand is given by $M_{it} = s_t(K_{it}, \omega_{it}^{L}, \omega_{it}^{H}, \bar{p}_{it})$, and it is strictly increasing in ω_{it}^{H}.*

This assumption generalizes Assumption 2.5 from the main model by allowing materials demand to depend on both input prices. In addition to these new assumptions, I retain Assumptions 2.1, 2.2, and 2.4 from the main model and state the following proposition.

Proposition OA-1.

(i) *Suppose Assumptions 2.1(i-ii) and 2.2 hold. With firm-level heterogeneity in input prices, the flexible input ratio, denoted by $\tilde{M}_{it} = M_{it}/L_{it}$, relies on K_{it}, ω_{it}^{L} and $p_{it}^{l/m}$:*

$$\tilde{M}_{it} = r_t(K_{it}, \omega_{it}^{L}, p_{it}^{l/m}). \tag{OA.1}$$

(ii) *Under Assumption 2.1(iii), $r_t(K_{it}, \omega_{it}^{L}, p_{it}^{l/m})$ is strictly monotone in ω_{it}^{L}.*

The proof is a straightforward extension of the proof of Proposition 2.1, and is therefore omitted. In contrast to Proposition 2.1, the flexible input ratio here also depends on the input price ratio $p_{it}^{l/m}$, implying that once we condition on $p_{it}^{l/m}$ and K_{it}, Equation (OA.1) is invertible to obtain ω_{it}^L.[43] By inverting Equation (OA.1) with respect to ω_{it}^L and then inverting the materials demand function in Assumption OA-2 with respect to ω_{it}^H, we can write productivity shocks as follows:

$$\omega_{it}^L = \bar{r}_t(K_{it}, \tilde{M}_{it}, p_{it}^{l/m}), \qquad \omega_{it}^H = \bar{s}_t(K_{it}, M_{it}, \tilde{M}_{it}, \bar{p}_{it}). \tag{OA.2}$$

The derivation of the control variables proceeds similarly to that in Section 3. First, I use Skorokhod's representation of ω_{it}^L, which gives:

$$\omega_{it}^L = g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1), \quad u_{it}^1 \mid \omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m} \sim \mathrm{U}(0,1). \tag{OA.3}$$

Unlike in Equation (3.1) of the main text, $g_1(\cdot)$ now includes both current and lagged input price ratios. This follows from Proposition OA-1, which indicates that the optimal input ratio depends on input price ratios. Combining Equations (OA.1), (OA.2), and (OA.3) gives:

$$\tilde{M}_{it} = r_t\big(K_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1), p_{it}^{l/m}\big) \equiv \tilde{r}_t(K_{it}, W_{it-1}, p_{it}^{l/m}, u_{it}^1), \tag{OA.4}$$

where $\tilde{r}_t(\cdot)$ is strictly monotone in u_{it}^1.

Lemma OA-4. *Under Assumptions 2.4, OA-1, and OA-2, u_{it}^1 is jointly independent of $(K_{it}, W_{it-1}, p_{it}^{l/m})$.*

Proof. This proof closely follows the proof of Lemma 3.1. By Assumption OA-1:

$$(p_{it}^{l/m}, \omega_{it}^L) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1},$$
$$p_{it}^{l/m}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}.$$

Monotonicity of $g_1(\cdot)$ with respect to its last argument implies that

$$u_{it}^1 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, p_{it}^{l/m}, p_{it-1}^{l/m}.$$

Since u_{it}^1 has a uniform distribution conditional on $(\omega_{it-1}^L, \omega_{it-1}^H, p_{it}^{l/m}, p_{it-1}^{l/m})$ by normalization in Equation (OA.3) and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$, we have that

$$u_{it}^1 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, p_{it}^{l/m}, p_{it-1}^{l/m} \sim \mathrm{U}(0,1).$$

Using Equation (OA.2), we substitute $(\omega_{it-1}^L, \omega_{it-1}^H)$ as functions of (W_{it-1}) to obtain:

$$u_{it}^1 \mid K_{it}, W_{it-1}, p_{it}^{l/m} \sim \mathrm{U}(0,1). \qquad \square$$

[43] Because of the cost function's properties, only the price ratio affects the flexible input ratio.

With this lemma and using Equation (OA.4), u_{it}^1 can be identified as:

$$u_{it}^1 = F_{\tilde{M}_{it}|K_{it},W_{it-1},p_{it}^{l/m}}(\tilde{M}_{it} \mid K_{it}, W_{it-1}, p_{it}^{l/m}).$$

Next, using Equations (OA.2) and (OA.3), we can write $\omega_{it}^L \equiv c_{1t}(W_{it-1}, p_{it}^{l/m}, u_{it}^1)$. Unlike in the main model, this control function includes the price ratio $p_{it}^{l/m}$ since u_{it}^1 is defined conditional on the price vector $\bar{p}_{it}$ in Equation (OA.3). The control function for ω_{it}^H follows a similar derivation:

$$\omega_{it}^H = g_2(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2), \qquad u_{it}^2 \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1 \sim \mathrm{U}(0,1).$$

Following the same steps used to derive Equation (3.2), the materials demand can be derived as $M_{it} \equiv \tilde{s}_t\big(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2\big)$, with $\tilde{s}_t(\cdot)$ strictly monotone in u_{it}^2.

Lemma OA-5. *Under Assumptions 2.4, OA-1 and OA-2, u_{it}^2 is jointly independent of $(K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1)$.*

Proof. This proof closely follows the proof of Lemma 3.2. By Assumption OA-1 we have

$$(\bar{p}_{it}, \omega_{it}^L, \omega_{it}^H) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1},$$
$$\bar{p}_{it}, g_1(\omega_{it-1}^L, \omega_{it-1}^H, p_{it-1}^{l/m}, p_{it}^{l/m}, u_{it}^1), g_2(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it-1}.$$

Monotonicity of $g_1(\cdot)$ and $g_2(\cdot)$ with respect to their last arguments and Lemma OA-4 imply that

$$u_{it}^2 \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it}, \bar{p}_{it-1}, u_{it}^1.$$

Since u_{it}^2 has a uniform distribution conditional on $(\omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it}, \bar{p}_{it-1}, u_{it}^1)$ and $(K_{it}, W_{it-1}) \in \mathcal{I}_{it-1}$ we have

$$u_{it}^2 \mid K_{it}, W_{it-1}, \omega_{it-1}^L, \omega_{it-1}^H, \bar{p}_{it}, u_{it}^1 \sim \mathrm{U}(0,1).$$

Using Equation (OA.2) to substitute $(\omega_{it-1}^L, \omega_{it-1}^H)$ as functions of W_{it-1}, I obtain

$$u_{it}^2 \mid K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1 \sim \mathrm{U}(0,1).$$ □

From this lemma and the strict monotonicity of M_{it} in u_{it}^2, we can recover u_{it}^2 as

$$u_{it}^2 = F_{M_{it}|K_{it},W_{it-1},\bar{p}_{it},u_{it}^1}(M_{it} \mid K_{it}, W_{it-1}, \bar{p}_{it}, u_{it}^1),$$

and the control function is given by $\omega_{it}^H \equiv c_{2t}(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2)$. Therefore, once we allow input prices to vary, the control functions take the form $\omega_{it}^L =$

$c_{1t}(W_{it-1}, p_{it}^{l/m}, u_{it}^1)$ and $\omega_{it}^H = c_{2t}(W_{it-1}, \bar{p}_{it}, u_{it}^1, u_{it}^2)$. The rest of the identification and estimation results remain the same with these modifications in the control variables.

B Models of Imperfect Competition

This section investigates the conditions under which the firm's materials demand function can be expressed as in Assumption 2.5, $M_{it} = s_t(K_{it}, \omega_{it}^H, \omega_{it}^L)$. To achieve this, I begin by deriving the firm's materials demand function from its cost minimization problem.

B.1 Symmetric and Aggregative Product Market Competition Games

In Equation (A.3) of Appendix A, the firm's static cost function is derived as follows:

$$C_t(\tilde{Y}_{it}, K_{it}, \omega_{it}^H, \omega_{it}^L, \bar{p}_t) = C_{1t}(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_{2t}(K_{it}, \bar{p}_t^l/\omega_{it}^L, p_t^m). \tag{OA.5}$$

where $\bar{p}_t = (p_t^l, p_t^m)$ is the input price vector. Here, $\tilde{Y}_{it}$ denotes the firm's static profit-maximizing level of output. By Shephard's Lemma, the firm's materials demand is given by:

$$M_{it} = \frac{\partial C_{1t}(K_{it}, \tilde{Y}_{it}, \omega_{it}^H) C_{2t}(K_{it}, \bar{p}_t^l/\omega_{it}^L, p_t^m)}{\partial p_t^m} = \tilde{s}_t(K_{it}, \tilde{Y}_{it}, \omega_{it}^H, \omega_{it}^L). \tag{OA.6}$$

Here, the firm's materials demand depends on K_{it}, target output $\tilde{Y}_{it}$, productivity shocks $(\omega_{it}^H, \omega_{it}^L)$, and the industry-specific input prices reflected in the t index in $\tilde{s}_t(\cdot)$. Next, compare this expression with the materials demand function given in Assumption 2.5.

$$M_{it} = s_t(K_{it}, \omega_{it}^H, \omega_{it}^L), \tag{OA.7}$$

Comparing Equation (OA.7) with Equation (OA.6) implies that, for our materials demand assumption to hold, the firm's profit-maximizing planned output $\tilde{Y}_{it}$ must take the following form: $\tilde{Y}_{it} = Y_t\big(K_{it}, \omega_{it}^H, \omega_{it}^L\big)$. In other words, the firm's profit-maximizing output choice should depend solely on $(K_{it}, \omega_{it}^H, \omega_{it}^L)$ and any industry-level variables which can be represented by the t index in $Y_t(\cdot)$. If this condition holds, the materials demand function derived from cost minimization in Equation

(OA.6) can be written as:

$$\begin{aligned} M_{it} &= \tilde{s}_t(K_{it}, Y_t\big(K_{it}, \omega_{it}^H, \omega_{it}^L\big), \omega_{it}^H, \omega_{it}^L) \\ &= s_t(K_{it}, \omega_{it}^H, \omega_{it}^L), \end{aligned}$$

which is consistent with Assumption 2.5. Hence, in the remainder of this section, I establish the sufficient conditions under which $\tilde{Y}_{it} = Y_t\big(K_{it}, \omega_{it}^H, \omega_{it}^L\big)$ holds.

To achieve this, I require two high-level conditions. First, there must be no unobserved, firm-specific demand shifters, as these would affect the firm's profit-maximizing output choice $\tilde{Y}_{it}$. Second, the firm's output choice must not be influenced by individual competitors' actions. Otherwise, $\tilde{Y}_{it}$ would depend on competitors' output (or price) decisions, which in turn depend on their productivity, contradicting $\tilde{Y}_{it} = Y_t(K_{it}, \omega_{it}^H, \omega_{it}^L)$. However, the output choice may still depend on common, industry-specific factors, which are captured by the t index in the $Y_t(\cdot)$ function.

One class of models satisfying these conditions is symmetric aggregative games, where a firm's best response depends only on industry-wide aggregates rather than on competitors' individual actions. When such industry-level aggregates can be constructed in best-response functions, the game can be represented as an aggregative game. Aggregative games are particularly useful because they reduce the dimensionality of the firm's strategic problem.

In what follows, I will provide conditions under which a firm's optimal output $\tilde{Y}_{it}$ depends on the aggregate in an aggregative game. I then present specific examples—such as Cournot and Bertrand competition (with logit and CES demand) in oligopolistic competition and CES/Kimball preferences in monopolistic competition—that can be framed as aggregative games.

Consider a market with J firms. Each firm produces a single product according to the cost function given in Equation (OA.5). Products may be either perfect or imperfect substitutes, depending on the assumed demand model. I assume each firm, in choosing $\tilde{Y}_{it}$, maximizes its static profit function given the strategic actions (price or quantity) of its rivals. A Nash equilibrium exists when no firm can unilaterally change its chosen action to achieve a higher profit, given all other firms' actions. To simplify the notation, let $\omega_i = (K_i, \omega_i^H, \omega_i^L)$ collect the cost determinants that firm i treats as given in static optimization problems. I also drop the t subscript throughout this section since the analysis here focuses on a single period. I start with two definitions.

Definition OA-1 (Aggregative Games, Acemoglu and Jensen 2013)**.** *A game is called (linearly) aggregative if each firm's payoff depends only on its own one-dimensional action, $a_i \geqslant 0$, and the sum of the actions of all firms, the aggregate, $A = \sum_{i=1}^{J} a_i$. Formally, $\pi_i(a_i, a_{-i}) = \pi_i(a_i, A)$, where a_{-i} denotes the action vector of all other firms.*

In our setting, assuming that a_i corresponds to firm i's quantity choice (for now, before considering a specific model), firm i's payoff can be expressed as

$$\pi_i = R_i(a) - C(a_i, \omega_i),$$

where $a = (a_1, a_2, \cdots, a_J)$ is the vector of actions across all firms, $R_i(\cdot)$ the revenue function of firm i and $C(a_i, \omega_i)$ is the cost function that is common to all firms in the industry since the production function is industry-specific.[44] If the game is aggregative, then the profit function becomes:

$$\begin{aligned}\pi_i(a_i, A, \omega_i) &= R_i(a_i, A) - C(a_i, \omega_i) \\ &= R_i(a_i, A_{-i} + a_i) - C(a_i, \omega_i),\end{aligned}$$

where $A_{-i} = \sum_{j \neq i} a_j$. In this formulation, the aggregative structure ensures that the revenue function depends only on the total actions of the rivals rather than each rival's individual decision. However, the revenue function can still be firm-specific (as indicated by index i) because it can be affected by firm-specific demand shocks. To address this, I also impose a symmetry condition:

Definition OA-2 (Symmetric Games)**.** *A competition game is symmetric (with respect to demand) if each firm's revenue satisfies $R_i(a_i, a_{-i}) = R_j(a_i, a_{-i})$ for any i, j.*

In other words, the firm identities are exchangeable: if two firms produce the same amount and face the same actions, their revenue is the same. In our setting, the symmetry condition implies that for any two firms i and j

$$R_i(a_i, A_{-i} + a_i) - C(a_i, \omega_i) = R_j(a_i, A_{-i} + a_i) - C(a_i, \omega_i).$$

[44] If a_i is the price, we can write quantity as a function of the aggregate and the firm's price, as I show in Online Appendix B.2.2 and Online Appendix B.2.3

Using this, we can write the profit function as:

$$\pi(a_i, A, \omega_i) = R(a_i, A_{-i} + a_i) - C(a_i, \omega_i)$$

Therefore, under symmetry, once cost determinants are included, the functional form for profit is no longer firm-specific (π_i vs π). In what follows, I write the profit function as $\pi(a_i, A, \omega_i)$. When the cost vector ω_i is suppressed, I use the shorter $\pi_i(a_i, A)$, where the subscript now signals the firm-specific cost variables.

With these assumptions, I aim to show that in a symmetric, aggregative competition game, (i) each firm's best response function depends only on the aggregate and its cost variables $a_i = \tilde{r}_i(A) = \tilde{r}(A, \omega_i)$ where $\tilde{r}_i(\cdot)$ is the best response function and (ii) an equilibrium exists and can be characterized by an aggregate A^* such that $a_i^* = \tilde{r}(A^*, \omega_i)$ and $\sum a_i^* = A^*$. The proof proceeds in three steps. First, I characterize the conditions under which a firm's best response can be expressed as a function of the aggregate (that is, A instead of A_{-i} because it is trivial to show that best response depends on A_{-i} with a well-behaved profit function). Second, I identify the conditions under which an equilibrium exists (that is, A^* exists). Finally, I give examples of some demand systems and competition games that satisfy these assumptions.

The analysis builds on Anderson et al. (2020), who study aggregative games in industrial organization to derive comparative statics of entry. For completeness, I reproduce some of their results and extend others to cover non-constant marginal cost cases.

Proposition OA-1 (Existence of Firm's Best Response as a function of the Aggregate)**.** *Assume that*

(i) The competition game is aggregative and symmetric.

(ii) $\pi_i(a_i, a_{-i})$ *is twice differentiable, and strictly quasi-concave in* a_i*, with a strictly negative second derivative with respect to* a_i *at an interior maximum.*

(iii) $\frac{d^2\pi_i}{da_i^2} < \frac{d^2\pi_i}{da_i dA_{-i}}$.

Then, each firm's best response can be written as a function of the sum of all firms' actions and its cost variables ω_i*. That is, the best response does not depend on the actions of the individual competitors.*

Proof. Since the game is aggregative, we can express the firm's profit function as

$$\pi_i(a_i, a_{-i}) = \pi_i(A_{-i} + a_i, a_i) = \pi(A_{-i} + a_i, a_i, \omega_i).$$

By the strict quasi-concavity specified in condition (ii), there exists a unique $a_i = r_i(A_{-i})$ that maximizes the firm's profit. I next show that $da_i/dA_{-i} = r_i'(A_{-i}) > -1$. Differentiating the FOC with respect to A_{-i} and applying the chain rule:

$$\frac{\partial^2 \pi_i}{\partial a_i^2}\frac{da_i}{dA_{-i}} + \frac{\partial^2 \pi_i}{\partial a_i \partial A_{-i}} = 0, \qquad \Longrightarrow \qquad r_i'(A_{-i}) = \frac{-\dfrac{\partial^2 \pi_i}{\partial a_i \partial A_{-i}}}{\dfrac{\partial^2 \pi_i}{\partial a_i^2}} > -1,$$

by condition (iii). Next, I show that there is a one-to-one mapping between A_{-i} and A. Differentiating the aggregate function $A = A_{-i} + r_i(A_{-i})$ with respect to A_{-i}:

$$\frac{dA}{dA_{-i}} = 1 + r_i'(A_{-i}) > 0,$$

because $r_i'(A_{-i}) > -1$. Hence, the mapping A_{-i} to A is strictly increasing and thus invertible, implying the existence of a function m_i such that:

$$A_{-i} = m_i(A).$$

Substituting $A_{-i} = m_i(A)$ into the best response function $r_i(\cdot)$ yields

$$a_i = r_i(A_{-i}) = r_i(m_i(A)) \equiv \tilde{r}_i(A) = \tilde{r}(A, \omega_i).$$

Therefore, each firm's optimal action can be expressed as a function of the aggregate and its cost variables, which concludes the proof. □

Next, I investigate the conditions that ensure the existence of an equilibrium. In particular, I show that there exists a fixed point A^* such that $\sum \tilde{r}_i(A^*) = A^*$.

Lemma OA-6 (Existence of Aggregate using Brouwer Fixed Point Theorem)**.** *Assume that each firm's strategy space is compact. Then there exists an A^* such that $\sum a_i = \sum \tilde{r}_i(A^*) = A^*$.*

Proof. An equilibrium exists if and only if the best response function as an aggregate $\tilde{r}_i(A)$ has a fixed point. This implies that the aggregate A satisfies:

$$A = \sum \tilde{r}_i(A) \equiv \bar{r}(A). \tag{OA.8}$$

To show the existence of a fixed point, observe that each individual best response function $\tilde{r}_i(A)$ is continuous. Since each firm's strategy space a_i lies within a compact interval $[a_i^{\min}, a_i^{\max}]$, the aggregate A must lie in the interval $\left[\sum a_i^{\min}, \sum a_i^{\max}\right]$, which is also compact. Hence, the aggregate best response function $\bar{r}(A)$ maps the compact interval $[A^{\min}, A^{\max}]$ into itself. Since $\bar{r}(A)$ is continuous and its domain is a compact convex set, the Brouwer Fixed Point Theorem guarantees the existence of a fixed point A^*. □

Lemma OA-7 (Existence of Aggregate using Intermediate Value Theorem)**.** *Assume that actions are strategic substitutes, that is, $r_i'(A_{-i}) < 0$. Then, there exists a unique A^* such that $\sum a_i = \sum \tilde{r}_i(A^*) = A^*$.*

Proof. First, I show that $d\tilde{r}_i(A)/dA < 1$. Differentiate $\tilde{r}_i(A)$ with respect to A:

$$\frac{d\tilde{r}_i(A)}{dA} = \frac{d}{dA} r_i(m_i(A)) = r_i'(m_i(A))m_i'(A). \tag{OA.9}$$

Since $m_i(\cdot)$ is a function of $A_i = A_{-i} + r_i(A_{-i})$, by the inverse function theorem:

$$m_i'(A) = \frac{1}{\frac{dA_i}{dA_{-i}}} = \frac{1}{1 + r_i'(m_i(A))}.$$

Substituting $m_i'(A)$ back into Equation (OA.9), we obtain:

$$\frac{d\tilde{r}_i(A)}{dA} = r_i'(m_i(A)) \cdot \frac{1}{1 + r_i'(m_i(A))}.$$

Since $0 > r_i'(A_{-i}) > -1$, this equation implies that $\tilde{r}_i'(A){<}0$, so $\tilde{r}_i(\cdot)$ is strictly decreasing.

The equilibrium condition can then be written as: $A = \bar{r}(A)$ with $\bar{r}(A) = \sum \tilde{r}_i(A)$. Since each $\tilde{r}_i(A)$ is strictly decreasing in A, their sum $\bar{r}(A)$ inherits this property and is strictly decreasing. Let $G(A) = \bar{r}(A) - A$. Given that $\bar{r}(A)$ is strictly decreasing, $G(A)$ is also strictly decreasing. I now impose the following boundary conditions:

$$\lim_{A \to 0} G(A) = \bar{r}(A) - A > 0, \qquad \lim_{A \to \infty} G(A) = \lim_{A \to \bar{A}} \bar{r}(A) - A < 0.$$

where $\bar{A}$ is the sum of the upper bound of the firms' strategy spaces. These boundary conditions hold under mild assumptions, such as bounded firm profits. Because $G(A)$ is strictly decreasing and meets these boundary conditions, it can cross zero at most once. Thus, $G(A) = 0$ has at most one solution. Consequently, there is at most one equilibrium A^* satisfying $\sum \tilde{r}_i(A^*) = A^*$. □

B.2 Examples of Aggregative Oligopolistic Competition Models

In this section, I demonstrate that under some assumptions, both Cournot competition and Nash–Bertrand competition with CES or logit demand meet the conditions of Proposition OA-1 and are, therefore, aggregative games.

B.2.1 Cournot Competition

In the Cournot model, the market price depends on the total quantity produced by the industry. Let $p(Q)$ be the inverse demand curve, where Q is the total output in the market. For firm i, let Q_{-i} be the total output of all other firms, so $Q = Q_{-i} + q_i$. The firm's profit function is given by

$$\pi_i(q) = \pi_i(Q, q_i) = p(Q)q_i - C_i(q_i),$$

where $C_i(q_i) = C(q_i, \omega_i)$. Redefining the strategic variable $a_i := q_i$ and $A_{-i} := Q_{-i}$, we can rewrite the profit function as:

$$\pi_i(a) = \pi_i(A, a_i) = p(A_{-i} + a_i)a_i - C_i(a_i).$$

Hence, Cournot competition is a symmetric aggregative game where the aggregate variable is the total industry quantity Q. This shows that condition (i) of Proposition OA-1 is satisfied. Next, I identify assumptions under which conditions (ii) and (iii) in Proposition OA-1 hold for the Cournot game.

Proposition OA-2. *Under the following assumptions, the Cournot competition game satisfies conditions (ii) and (iii) of Proposition OA-1:*

(i) The cost function is increasing and convex, demand is downward sloping: $C_i'(a_i) > 0$, $C_i''(a_i) \geqslant 0$ *for all* i *and* $p'(A) < 0$.

(ii) The marginal revenue is decreasing: $2p'(A) + a_i p''(A) < 0$ *for all* i.

Proof. I first show that the profit function is strictly quasi-concave by examining its second derivative

$$\frac{d^2\pi_i}{da_i^2} = \big(2p'(A) + a_i\, p''(A)\big) - C_i''(a_i).$$

By condition (ii), the first term is negative, and by condition (i), the second term is also negative. Hence, the profit function is concave, satisfying the condition (ii) of Proposition OA-1. Next, I show that the condition (iii) of Proposition OA-1 also holds: $\frac{d^2\pi_i}{da_i^2} < \frac{d^2\pi_i}{da_i dA_{-i}}$. In the Cournot model, this difference can be written as:

$$\begin{aligned}\frac{d^2\pi_i}{da_i^2} - \frac{d^2\pi_i}{da_i\, dA_{-i}} &= (2\,p'(A) + a_i\,p''(A) - C_i''(a_i)) - (p'(A) + a_i\,p''(A)) \\ &= p'(A) - C_i''(a_i) < 0,\end{aligned}$$

since $p'(A) < 0$ and $C_i''(a_i) \geqslant 0$ by Assumption (i). This concludes that the firm's best response can be expressed as a function of aggregate $a_i = r_i(A)$. □

To show that an equilibrium A^* exists, I impose the trivial assumption that each firm's output is restricted to a finite interval, $a_i \in [0, \bar{a}]$. This makes the strategy space compact and by Lemma OA-6, there exists A^* such that $\sum_i r_i(A^*) = A^*$, ensuring the existence of an equilibrium.

B.2.2 Bertrand Competition with Logit Demand

In a symmetric Bertrand oligopoly game with logit demand, firm i's residual demand is given by:

$$D(p_i, p_{-i}) = \frac{\exp[(s - p_i)/\mu]}{\sum_j \exp[(s - p_j)/\mu]},$$

where s is the "quality" parameter, p_i is price of firm i, and $\mu > 0$ represents the degree of preference heterogeneity. Let $a_i = \exp[(s - p_i)/\mu]$. Then, we can transform the demand function to write it as a function of a_i as follows:

$$D(a_i, A) = \frac{a_i}{A}, \quad \text{with} \quad A = \sum_j \exp[(s - p_j)/\mu].$$

Using this notation, the firm's profit function becomes:

$$\pi_i = (s - \mu \ln a_i)q_i - C_i(q_i), \qquad q_i := D(a_i, A) = \frac{a_i}{A_{-i} + a_i}.$$

Hence, the Nash–Bertrand game with logit demand is an aggregative game with respect to the variable a_i, satisfying condition (i) of Proposition OA-1. I now present the assumptions under which conditions (ii) and (iii) of Proposition OA-1 are also satisfied.

Lemma OA-8. *Under the following assumptions, the Bertrand competition with logit demand satisfies conditions (ii) and (iii) of Proposition OA-1:*

(i) The cost function is convex and increasing: $C_i''(\cdot) \geqslant 0$, $C_i'(\cdot) > 0$ *for all* i.

(ii) Positive markup: $p_i - C_i'(\cdot) > 0$ for all i in equilibrium.

Proof. I begin by deriving the derivatives of the residual demand $q_i = D(a_i, A)$ with respect to A_{-i} and a_i

$$\frac{\partial q_i}{\partial a_i} = \frac{A_{-i}}{(A_{-i}+a_i)^2} > 0, \qquad \frac{\partial q_i}{\partial a_i \partial a_i} = -\frac{2A_{-i}}{(A_{-i}+a_i)^3} < 0, \tag{OA.10}$$

$$\frac{\partial^2 q_i}{\partial a_i \partial A_{-i}} = \frac{a_i - A_{-i}}{(A_{-i}+a_i)^3}, \qquad \frac{\partial q_i}{\partial A_{-i}} = -\frac{a_i}{(A_{-i}+a_i)^2} < 0 \tag{OA.11}$$

Next, I compute the derivatives of the profit function as follows:

$$\frac{d^2\pi_i}{da_i^2} = \frac{\mu q_i}{a_i^2} - \frac{2\mu}{a_i}\frac{dq_i}{da_i} + (s - \mu\log(a_i) - C_i'(q_i))\frac{d^2 q_i}{da_i^2} - C_i''(q_i)\left(\frac{dq_i}{da_i}\right)^2$$

$$\frac{\partial^2\pi_i}{\partial a_i \partial A_{-i}} = -\frac{\mu}{a_i}\frac{\partial q_i}{\partial A_{-i}} - C_i''(q_i)\frac{\partial q_i}{\partial a_i}\frac{\partial q_i}{\partial A_{-i}} + (s - \mu\log(a_i) - C_i'(q_i))\frac{\partial^2 q_i}{\partial a_i \partial A_{-i}}$$

The difference between these two expressions can be written as:

$$\begin{aligned}\frac{d^2\pi_i}{da_i^2} - \frac{d^2\pi_i}{da_i dA_{-i}} = {} & \frac{\mu}{a_i}\left(\frac{q_i}{a_i} - 2\frac{\partial q_i}{\partial a_i} + \frac{\partial q_i}{\partial A_{-i}}\right) + (s - \mu\log(a_i) - C_i'(q_i))\left(\frac{\partial^2 q_i}{\partial a_i^2} - \frac{\partial^2 q_i}{\partial a_i \partial A_{-i}}\right) \\ & - C_i''(q_i)\frac{\partial q_i}{\partial a_i}\left(\frac{\partial q_i}{\partial a_i} - \frac{\partial q_i}{\partial A_{-i}}\right).\end{aligned}$$

Substituting the derivatives of q_i in Equations (OA.10-OA.11), we obtain

$$\begin{aligned}\frac{d^2\pi_i}{da_i^2} - \frac{d^2\pi_i}{da_i dA_{-i}} = {} & \underbrace{\frac{\mu}{a_i}\left(-\frac{A_{-i}}{(A_{-i}+a_i)^2}\right)}_{(1)} + \underbrace{(s - \mu\log(a_i) - C_i'(q_i))\left(-\frac{1}{(A_{-i}+a_i)^2}\right)}_{(2)} \\ & \underbrace{- C_i''(q_i)\left(\frac{A_{-i}}{(A_{-i}+a_i)^2}\right)\left(\frac{1}{A_{-i}+a_i}\right)}_{(3)}.\end{aligned}$$

Observe that $(s - \mu\log(a_i) - C_i'(q_i))$ equals $p_i - C_i'(q_i)$, which is positive by condition (ii). Consequently, term (2) is negative. Term (3) is weakly negative due to the cost function's convexity (condition (i)). Since term (1) is also negative, the entire expression is negative, thus establishing that condition (ii) of Proposition OA-1 is satisfied.

I now show that the firm's profit function is strictly quasi-concave in a_i. To that end, I first express the profit function in terms of q_i, substituting a_i as a function of q_i:

$$\pi_i(q_i) = \underbrace{\big(s - \mu[\ln q_i + \ln A_{-i} - \ln(1-q_i)]\big)q_i}_{4} \underbrace{-C_i(q_i)}_{5}.$$

It is straightforward to verify that term (4) is concave in q_i, and term (5) is also concave since the cost function is convex. Hence, the profit function is quasi-concave in q_i. Furthermore, q_i is a monotone in a_i. Since a monotone transformation of a quasi-concave function remains quasi-concave, the profit function is quasi-concave in a_i. This proves that the condition (iii) of Proposition OA-1 is satisfied. $\square$

To show that an equilibrium A^* exists, I cannot rely on Brouwer's Fixed Point Theorem in Lemma OA-6 because the profit function is infinite at $a_i = 0$. Instead, I apply Lemma OA-7. To apply this Lemma, note that the firm's best response p_i is decreasing in p_j with $j \neq i$ as prices are strategic substitutes in the Bertrand competition game. This implies that a_i's are also strategic substitutes since they are monotone transformations of p_i's. Therefore, Lemma OA-7 guarantees the existence of an equilibrium.

B.2.3 Bertrand Competition with CES Demand

In the CES demand system, firm i's demand is given by

$$D_i(p) = \frac{p_i^{-\lambda-1}}{\sum p_j^{-\lambda}},$$

with $\lambda = \frac{\rho}{1-\rho}$ and $\rho \in (0,1)$ is the substitution parameter. Let $a_i \equiv p_i^{-\lambda}$. Then, the demand for firm i can be expressed as

$$D_i(a) = \frac{a_i^{(\lambda+1)/\lambda}}{\sum_{j\neq i} a_j + a_i} = \frac{a_i^{(\lambda+1)/\lambda}}{A_{-i} + a_i},$$

where $A_{-i} \equiv \sum_{j\neq i} a_j$. Hence, the firm's profit function becomes:

$$\pi_i = (a_i^{-1/\lambda})q_i - C_i(q_i), \qquad q_i = \frac{a_i^{(\lambda+1)/\lambda}}{A_{-i} + a_i}.$$

We thus observe that Bertrand competition under CES demand constitutes a symmetric aggregative game in the variable a_i, satisfying condition (i) of Proposition OA-1. Next, we show that under some assumptions, conditions (ii) and (iii) of Proposition OA-1 are also satisfied.

Lemma OA-9. *Under the assumptions below, the Bertrand competition with CES demand satisfies conditions (ii) and (iii) of Proposition OA-1.*

(i) The cost function is convex and increasing: $C_i''(q_i) \geqslant 0$, $C_i'(q_i) > 0$.

(ii) Substitution parameter ρ *is less than 0.5.*

(iii) Positive markup: $p_i - C_i'(\cdot) > 0$ *for all* i *in equilibrium.*

Proof. Note that the profit function can be written as:

$$\pi_i(A_{-i}, a_i) = \frac{a_i}{A_{-i} + a_i} - C_i(q_i).$$

First, we show that $\frac{d^2\pi_i}{da_i^2} - \frac{d^2\pi_i}{da_i\, dA_{-i}} < 0$. We begin by computing the necessary derivatives:

$$\frac{\partial^2 \pi_i}{\partial a_i\, \partial A_{-i}} = \frac{a_i - A_{-i}}{(A_{-i}+a_i)^3} - C_i''(q_i)\left(\frac{\partial q_i}{\partial a_i}\right)\left(\frac{\partial q_i}{\partial A_{-i}}\right) - C_i'(q_i)\frac{\partial^2 q_i}{\partial a_i\, \partial A_{-i}}$$

$$\frac{d^2\pi_i}{da_i^2} = -\frac{2A_{-i}}{(A_{-i}+a_i)^3} - C_i'(q_i)\frac{\partial^2 q_i}{\partial a_i^2} - C_i''(q_i)\left(\frac{\partial q_i}{\partial a_i}\right)^2.$$

The difference can be written as

$$\frac{d^2\pi_i}{da_i^2} - \frac{d^2\pi_i}{da_i, dA_{-i}} = \underbrace{C_i'(q_i)\left(-\frac{\partial^2 q_i}{\partial a_i^2} + \frac{\partial^2 q_i}{\partial a_i \partial A_{-i}}\right)}_{1} + \underbrace{C_i''(q_i)\cdot \frac{\partial q_i}{\partial a_i}\left(-\frac{\partial q_i}{\partial a_i} + \frac{\partial q_i}{\partial A_{-i}}\right)}_{2} - \underbrace{\frac{1}{(A_{-i}+a_i)^2}}_{3},$$

(2) is negative because $C_i''(q_i) > 0$ and

$$\frac{dq_i}{da_i} = \frac{a_i^{\gamma-1}\gamma A_{-i} + (\gamma-1)a_i^{\gamma}}{(A_{-i}+a_i)^2} > 0, \qquad \frac{\partial q_i}{\partial A_{-i}} = \frac{a_i^{\gamma}}{(A_{-i}+a_i)^2} < 0,$$

where γ is defined as $\gamma := (\lambda+1)/\lambda$. Hence, we need to show that (1) + (3) is negative. We assume the opposite to derive a contradiction. Focusing on term (1)

$$(1) = C_i'(q_i)\Big(-\frac{\partial^2 q_i}{\partial a_i^2} + \frac{\partial^2 q_i}{\partial a_i \partial A_{-i}}\Big) = -C_i'(q_i)\gamma\frac{(\gamma-1)A_{-i} + (\gamma-2)a_i}{(A_{-i}+a_i)^2}a_i^{\gamma-2},$$

Combining terms (1) and (3) yields

$$(1) + (3) = -\gamma\frac{(\gamma-1)A_{-i} + (\gamma-2)a_i}{(A_{-i}+a_i)^2}a_i^{\gamma-2}C_i'(q_i) - \frac{1}{(A_{-i}+a_i)^2} \geqslant 0,$$

which in turn implies that

$$-\gamma\big((\gamma-1)A_{-i} + \gamma(\gamma-2)a_i\big)a_i^{\gamma-2}C_i'(q_i) > 1. \tag{OA.12}$$

By substituting $p_i = a_i^{1-\gamma}$, the condition in Equation (OA.12) can be written as:

$$-\gamma\big((\gamma-1)A_{-i} + \gamma(\gamma-2)a_i\big)\frac{1}{a_i}\frac{C_i'(q_i)}{p_i} > 1 \quad\implies\quad -\gamma\big((\gamma-1)A_{-i} + \gamma(\gamma-2)a_i\big)\frac{1}{a_i} > \frac{p_i}{C_i'(q_i)}$$

Simplifying further, we obtain

$$\underbrace{-\gamma(\gamma-1)(A_{-i}/a_i)}_{(4)}\ \underbrace{-\gamma(\gamma-2)}_{(5)} > \frac{p_i}{C_i'(q_i)}. \tag{OA.13}$$

Note that term (4) is negative and is maximized (i.e., equals 0) when $\gamma = 1$. Likewise, (5) is maximized at $\gamma = 1$ and equals 1. Hence, the entire left-hand side is bounded above by 1. By condition (iii) $p_i/C_i'(q_i) \geq 1$, therefore Equation

(OA.13) cannot be satisfied, giving a contradiction. This concludes that condition (ii) of Proposition OA-1 holds.

We now show that the profit function is quasi-concave in a_i. The second derivative of the profit function is:

$$\frac{d^2\pi_i}{da_i^2} = \underbrace{-\frac{2A_{-i}}{(A_{-i}+a_i)^3}}_{(1)} \underbrace{-C_i'(q_i)\frac{\partial^2 q_i}{\partial a_i^2}}_{(2)} \underbrace{-C_i''(q_i)\left(\frac{\partial q_i}{\partial a_i}\right)^2}_{(3)}.$$

Because terms (1) and (3) are negative, we only need (2) to be negative. Given that $C_i'(q_i)$ is positive, this reduces to showing $\partial^2 q_i/\partial a_i^2$ is positive:

$$\frac{\partial^2 q_i}{\partial a_i^2} = \frac{a_i^{\gamma-1}\left[\frac{\gamma(\gamma-1)A_{-i}^2}{a_i} + 2\gamma(\gamma-2)A_{-i} + (\gamma-1)(\gamma-2)a_i\right]}{(A_{-i}+a_i)^3}.$$

Note that condition (ii) implies that $\gamma > 2$. Therefore, the above expression is positive. It follows that the profit function is concave in a_i. □

Finally, to establish the existence of an equilibrium, we apply Lemma OA-7 because prices, and therefore a_i's, are strategic substitutes in Nash-Bertrand competition with CES demand.

B.3 Examples of Aggregative Monopolistic Competition Models

Monopolistic competition describes a market structure where firms are atomistic, and their decisions do not affect market-level variables such as prices and quantities. Under certain conditions, monopolistic competition models can be formulated as an aggregative game in which each firm's profit is expressed in terms of market-level aggregates and the firm's strategic choice. We begin by stating these conditions. Let i index a firm and assume there is a continuum of firms $i \in [0, N]$. Let u be the utility function that represents consumer preferences over the firms' products, and let E be the total budget.

Assumption OA-1 (Additive Separability)**.** *Preferences are additively separable*

$$\max_{\{q_i\}_{i\in[0,N]}} \int_0^N u_i(q_i)\,di \quad \textit{subject to} \quad \int_0^N p_i q_i\,di = E,$$

Assumption OA-2 (Symmetric Preferences)**.** *u is* symmetric *over* $[0, N]$ *that is for any Lebesgue measure-preserving mapping* $\tau : [0, N] \to [0, N]$ $u(\{q_i\}_{i\in[0,N]}) = u(\{q_{\tau(i)}\}_{i\in[0,N]})$.

The assumptions of additive separability and symmetry imply that the consumer's utility maximization problem can be written as:

$$\max_{\{q_i\}_{i\in[0,N]}} \int_0^N u(q_i)\, di \quad \text{subject to} \quad \int_0^N p_i q_i \, di = E.$$

Solving the FOC, we obtain the following inverse demand:

$$p(q_i) = \frac{u'(q_i)}{\lambda} \quad \text{with} \quad \lambda = \frac{\int_0^N q_i u'(q_i)\, di}{E}. \tag{OA.14}$$

From this inverse demand, the firm's profit function can be written as:

$$\pi_i(q_i, \omega_i, \lambda) = \frac{u'(q_i) q_i}{\lambda} - C(q_i, \omega_i),$$

where λ is an aggregate term that depends on the distribution of q_i. This profit function satisfies the definition of an aggregative game given in Definition OA-1. Hence, a monopolistic competition model that satisfies additive separability and symmetry can be represented as an aggregative game. Solving the FOC of firm i's profit maximization problem:

$$u''(q_i^*) q_i^* + u'(q_i^*) = \lambda\, C_i'(q_i^*). \tag{OA.15}$$

where $C_i(q_i^*) = C(q_i^*, \omega_i)$. This equation implicitly defines the firm's optimal quantity choice as a function of λ and its own cost parameters ω_i. Having derived the best response $q_i^*(\lambda, \omega_i)$, we substitute it into the definition of λ in Equation (OA.14) to obtain:

$$\lambda = \frac{\int_0^N q_i^*(\lambda, \omega_i) u'(q_i^*(\lambda, \omega_i))\, di}{E}. \tag{OA.16}$$

This equation defines a fixed point relationship with respect to λ, thus characterizing the equilibrium. Next, I present several commonly used monopolistic competition models under this family following Arkolakis and Morlacco (2017) and Parenti et al. (2017).

B.3.1 CES Preferences

Following Melitz (2003), the CES demand system has been widely employed in models of monopolistic competition. The preferences are given by:

$$U = \left(\int_0^N q_i^\rho di \right)^{1/\rho}.$$

Solving this system yields the following demand function for firm i:

$$p_i(q_i, P, Q) = P\left(\frac{q_i}{Q}\right)^{-1/\sigma} \quad \text{with} \quad Q = \left(\int_0^N q_i^\rho \, di\right)^{1/\rho} \quad P = \left(\int_0^N p_i^{1-\sigma} \, di\right)^{1/(1-\sigma)},$$

where $\sigma = 1/(1-\rho)$. Using this inverse demand, we write the profit function as:

$$\pi_i(q_i, \omega_i, P, Q) = P\left(\frac{q_i}{Q}\right)^{-1/\sigma} q_i - C(q_i, \omega_i).$$

Therefore, monopolistic competition with CES preferences is an aggregative game. However, one drawback of CES preferences is that they imply a constant markup. The following demand systems allow for markups that vary across firms.

B.3.2 Kimball Preferences

Kimball preferences, used by Klenow and Willis (2016), are implicitly defined by the following condition:

$$\min \int_0^N p_i q_i di \qquad \text{s.t} \qquad \int_0^N Y\Big(\frac{q_i}{Q}\Big) di = 1,$$

where $Q = \int_0^N q_i di$ and $Y(\cdot)$ is a function such that $Y(1) = 1$, $Y' > 0$ and $Y''(\cdot) < 0$. One can derive the demand function by solving a standard cost minimization problem in which prices and the aggregate quantity Q are given. Let Z denote the inverse of the derivative of the function $Y(\cdot)$. Then, the demand function for firm i is given by:

$$q(p_i, Q, \tilde{Q}, P) = Z\Big(\tilde{Q}\frac{p_i}{P}\Big)Q,$$

with $P = \int_0^N p_i \frac{q_i}{Q} di$ and $\tilde{Q} = \int_0^N Y'\Big(\frac{q_i}{Q}\Big)\frac{q_i}{Q} di$. The profit function is given by

$$\pi(p_i, \omega_i, \lambda) = q(p_i, Q, \tilde{Q}, P)p_i - C(q(p_i, Q, \tilde{Q}, P), \omega_i)$$

which depends on three aggregates, Q, P and $\tilde{Q}$, and the cost parameters of firm i. Hence, Kimball preferences also yield an aggregative game.

B.3.3 Quadratic Preferences

Quadratic preferences, used by Melitz and Ottaviano (2008), are defined by the following preferences:

$$U(q) = \alpha \int_0^N q_i di - \frac{\gamma}{2} \int_0^N q_i^2 di - \frac{\eta}{2}\Big(\int_0^N q_i^2 di\Big)^2.$$

Here, α and η govern substitution with the numeraire, while γ measures product differentiation. The inverse demand for each product i is then given by

$$p_i(q_i, Q) = \alpha - \gamma q_i - \eta Q,$$

where $Q = \int_0^N q_i di$ is the aggregate. Using this inverse demand, we can express the profit function as:

$$\pi(q_i, \omega_i, Q) = (\alpha - \gamma q_i - \eta Q)q_i - C(q_i, \omega_i),$$

which satisfies the properties of an aggregative game.

B.3.4 Stone-Geary Preferences

In the Stone-Geary demand system, the corresponding utility function is given by:

$$U = \int_0^N \alpha \ln(q_i - \gamma)\, di.$$

Maximizing this utility subject to the budget constraints yields the following inverse demand function for firm i:

$$p_i(q_i, E, G) = \frac{\alpha(E - G)}{(q_i - \gamma)} \quad \text{with} \quad E = \int_0^N p_i q_i\, di \quad \text{and} \quad G = \int_0^N p_i \gamma\, di.$$

From this inverse demand, the firm's profit function can be written as:

$$\pi_i(q_i, \omega_i, E, G) = \frac{\alpha(E - G)}{(q_i - \gamma)} q_i - C(q_i, \omega_i).$$

Thus, monopolistic competition with Stone-Geary preferences is aggregative.

C Application to Parametric Production Functions

This section applies the identification strategy to specific functional forms. My model nests five labor-augmenting production functions:

$$y_{it} = f_t\big(K_{it}, h_t(K_{it}, \omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it} \qquad \text{(Weak Homot. Sep.)}$$

$$y_{it} = f_t\big(K_{it}, h_t(\omega_{it}^L L_{it}, M_{it})\big) + \omega_{it}^H + \epsilon_{it} \qquad \text{(Strong Homot. Sep.)}$$

$$y_{it} = v k_{it} + f_t\big(\tilde{L}_{it} h_t(\omega_{it}^L, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it} \qquad \text{(Homogeneous)}$$

$$y_{it} = \frac{v}{\sigma} \log\Big(\beta_k K_{it}^\sigma + (1 - \beta_k)\big(\beta_l \big[\omega_{it}^L L_{it}\big]^{\sigma_1} + (1 - \beta_l) M_{it}^{\sigma_1}\big)^{\frac{\sigma}{\sigma_1}}\Big) + \omega_{it}^H + \epsilon_{it} \qquad \text{(Nest. CES)}$$

$$y_{it} = \frac{v}{\sigma} \log\Big((1 - \beta_l - \beta_m) K_{it}^\sigma + \beta_l \big(\omega_{it}^L L_{it}\big)^\sigma + \beta_m M_{it}^\sigma\Big) + \omega_{it}^H + \epsilon_{it} \qquad \text{(CES)}$$

where $\tilde{L}_{it} = L_{it}/K_{it}$ in the homogeneous production function, which is obtained by dividing the arguments of the production function by K_{it}. I first illustrate how to estimate CES, Nested CES, and homogeneous production functions with labor-augmenting productivity. Then, I show the application of Imbens and Newey (2009) to the Cobb-Douglas model under Hicks-Neutral productivity. For simplicity, I omit the time subscripts from the production function parameters. I maintain the assumptions in Section 2.2 except the form of the production function.

C.1 Nested CES Production Function

We write the log of the Nested CES production function as:

$$y_{it} = \frac{v}{\sigma}\log\Big(\beta_k K_{it}^{\sigma} + (1-\beta_k)\left(\beta_l\left[\omega_{it}^{L} L_{it}\right]^{\sigma_1} + (1-\beta_l)M_{it}^{\sigma_1}\right)^{\sigma/\sigma_1}\Big) + \omega_{it}^{H} + \epsilon_{it}.$$

By using its homotheticity, we can write the Nested CES production function as:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left(\beta_k \tilde{K}_{it}^{\sigma} + (1-\beta_k)\big(\beta_l\left[\omega_{it}^{L}\tilde{L}_{it}\right]^{\sigma_1} + (1-\beta_l)\big)^{\sigma/\sigma_1}\right) + \omega_{it}^{H} + \epsilon_{it},$$

where $\tilde{K}_{it} := K_{it}/M_{it}$ and $\tilde{L}_{it} := L_{it}/M_{it}$ and $m_{it} := \log(M_{it})$. From the FOCs of cost minimization we have $\omega_{it}^{L} = \gamma\tilde{L}_{it}^{(1-\sigma_1)/\sigma_1}$, where $\gamma = \big((1-\beta_l)p_t^l/(\beta_l p_t^m)\big)^{1/\sigma_1}$ is a constant. Substituting this expression into the production function gives

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left(\beta_k \tilde{K}_{it}^{\sigma} + (1-\beta_k)\gamma_1\big(\tilde{L}_{it} + \gamma_2\big)^{\sigma/\sigma_1}\right) + \omega_{it}^{H} + \epsilon_{it},$$

where $\gamma_1 = \beta_l\gamma^{\sigma_1}$ and $\gamma_2 = (1-\beta_l)/(\beta_l\gamma^{\sigma_1})$. This production function can be estimated by replacing ω_{it}^{H} with its control function, which leads to the following estimating equation:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left(\beta_k \tilde{K}_{it}^{\sigma} + (1-\beta_k)\gamma_1\big(\tilde{L}_{it} + \gamma_2\big)^{\sigma/\sigma_1}\right) + c_{2t}(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it}.$$

In this equation, the levels of β_k and $(1-\beta_k)\gamma_1$ cannot be separately identified from $c_{2t}(\cdot)$ because scaling these parameters by the same value becomes an additive constant (and hence part of $c_{2t}(\cdot)$) in a log production function. However, their ratio is identified. Next, we write the elasticities of the flexible input and capital as:

$$\theta_{it}^{V} = v\frac{(1-\beta_k)\gamma_1 x_{it}^{\sigma}}{(1-\beta_k)\gamma_1 x_{it}^{\sigma} + \beta_k K_{it}^{\sigma}}, \qquad \theta_{it}^{K} = v\frac{\beta_k K_{it}^{\sigma}}{(1-\beta_k)\gamma_1 x_{it}^{\sigma} + \beta_k K_{it}^{\sigma}},$$

where $x_{it} = M_{it}(\tilde{L}_{it} + \gamma_2)^{1/\sigma_1}$. Note that θ_{it}^{V} and θ_{it}^{K} are identified using these expressions as they depend only on the ratio of β_k and $(1-\beta_k)\gamma_1$ rather than their individual values. This result is the parametric counterpart of Proposition 4.1, which shows that only the sum of labor and materials elasticities (θ_{it}^{V}) is identified from input and output data. However, labor and materials elasticities can still be obtained from θ_{it}^V using the ratio of revenue shares as in Equation (4.6).

C.2 CES Production Function

Using its homotheticity, we express the log of the CES production function as:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \beta_l\left[\omega_{it}^L\tilde{L}_{it}\right]^{\sigma} + \beta_m\right) + \omega_{it}^H + \epsilon_{it}.$$

The FOCs of cost minimization imply that $\omega_{it}^L = \gamma\tilde{L}_{it}^{(1-\sigma)/\sigma}$, where γ is a constant that depends on input prices and model parameters. Substituting this into the production function yields:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \gamma_1(\tilde{L}_{it}+\gamma_2)\right) + \omega_{it}^H + \epsilon_{it},$$

where $\gamma_1 := \gamma^{\sigma}\beta_l$ and $\gamma_2 := \beta_m/(\gamma^{\sigma}\beta_l)$. The model parameters can be estimated using the following equation:

$$y_{it} = vm_{it} + \frac{v}{\sigma}\log\left((1-\beta_l-\beta_m)\tilde{K}_{it}^{\sigma} + \gamma_1(\tilde{L}_{it}+\gamma_2)\right) + c_{2t}(W_{it-1}, u_{it}^1, u_{it}^2) + \epsilon_{it}.$$

As with the Nested CES model, we identify the sum of the flexible input elasticities and the capital elasticity from the model parameters as follows:

$$\theta_{it}^V = v\frac{\gamma_1 x_{it}^{\sigma}}{\gamma_1 x_{it}^{\sigma} + (1-\beta_l-\beta_m)K_{it}^{\sigma}}, \qquad \theta_{it}^K = v\frac{(1-\beta_l-\beta_m)K_{it}^{\sigma}}{\gamma_1 x_{it}^{\sigma} + (1-\beta_l-\beta_m)K_{it}^{\sigma}},$$

where $x_{it} = M_{it}(\tilde{L}_{it}+\gamma_2)$. These elasticities are identified because they depend only on the ratio of $(1-\beta_l-\beta_m)$ and γ_1. In turn, labor and materials elasticities can be recovered from θ_{it}^V using the revenue shares as in Equation (4.6).

C.3 Imposing Homogeneity Restriction on the Production Function

A production function with homogeneous of degree v can be written as:

$$y_{it} = vk_{it} + f_t\big(1, \tilde{L}_{it}h_t(\omega_{it}^L, \tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it},$$

where $\tilde{L}_{it} = L_{it}/K_{it}$. The reduced form of this production function is

$$y_{it} = vk_{it} + \tilde{f}_t\big(\tilde{L}_{it}\bar{h}_t(\tilde{M}_{it})\big) + \omega_{it}^H + \epsilon_{it},$$

where $\tilde{f}_t = f_t\big(1, \tilde{L}_{it}\bar{h}_t(\tilde{M}_{it})\big)$. The identification results in Sections 3 and 4 apply to this model once the homogeneity condition is imposed.

C.4 Application of Imbens and Newey (2009) to Cobb-Douglas Form

For illustration, I use a value-added production function under Hicks-neutral productivity, since Ackerberg et al. (2015) showed that the gross Cobb-Douglas form with two flexible inputs is not identified:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it}^H + \epsilon_{it}.$$

I use the standard assumptions from the proxy variable literature: (i) productivity follows an exogenous first-order Markov process $P(\omega_{it}^H \mid \mathcal{I}_{it-1}) = P(\omega_{it}^H \mid \omega_{it-1}^H)$, (ii) capital is a dynamic input while labor is a static input optimized each period, and (iii) the firm's intermediate input decision is given by $m_{it} = s(k_{it}, \omega_{it}^H)$, which is strictly increasing in ω_{it}^H. Using these assumptions, I construct a control variable following the steps outlined in Section 3 as follows:

$$\omega_{it}^H = g(\omega_{it-1}^H, u_{it}), \qquad u_{it} \mid \omega_{it-1}^H \sim \mathrm{U}(0,1). \tag{OA.17}$$

From the Markov Assumption, we have that $\omega_{it}^H \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^H$. Substituting ω_{it}^H using Equation (OA.17) we obtain $g(\omega_{it-1}^H, u_{it}) \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^H$, which implies that $u_{it} \perp\!\!\!\perp \mathcal{I}_{it-1} \mid \omega_{it-1}^H$. Using this,

$$m_{it} = s(k_{it}, \omega_{it}^H) = s(k_{it}, g(\omega_{it-1}^H, u_{it})) \equiv \tilde{s}(k_{it}, k_{it-1}, m_{it-1}, u_{it}).$$

Since $s(k_{it}, \omega_{it}^H)$ is strictly increasing in ω_{it}^H and $g(\omega_{it-1}^H, u_{it})$ is strictly increasing in u_{it}, $\tilde{s}(\cdot)$ is strictly increasing in u_{it}. It follows from Lemma 3.1 that

$$u_{it} \mid k_{it}, m_{it-1}, k_{it-1} \sim \mathrm{U}(0,1). \tag{OA.18}$$

Thus, we can recover u_{it} as the conditional CDF of m_{it}: $u_{it} = F_{m_{it}}(m_{it} \mid k_{it}, m_{it-1}, k_{it-1})$. In turn, this lets us construct a proxy for ω_{it}^H as a function of $(m_{it-1}, k_{it-1}, u_{it})$:

$$\omega_{it}^H = g(\omega_{it-1}^H, u_{it}) = g(s^{-1}(k_{it-1}, m_{it-1}), u_{it}) \equiv c_1(m_{it-1}, k_{it-1}, u_{it}).$$

Using this expression, we arrive at a partially linear model:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_1(m_{it-1}, k_{it-1}, u_{it}) + \epsilon_{it}. \tag{OA.19}$$

We can derive additional moment restrictions by exploiting the first-order Markov property of ω_{it}^H. Specifically, using $\omega_{it}^H = c_2(\omega_{it-1}^H) + \xi_{it}$ and $\mathbb{E}[\xi_{it} \mid \mathcal{I}_{it-1}] = 0$,

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + c_2(m_{it-1}, k_{it-1}) + \xi_{it} + \epsilon_{it}, \tag{OA.20}$$

with $\mathbb{E}[\xi_{it} \mid \mathcal{I}_{it-1}] = 0$. We can estimate the parameters (β_k, β_l) and functions $c_1(\cdot), c_2(\cdot)$ from Equations (OA.19) and (OA.20) by using the following moments:

$$\mathbb{E}[\epsilon_{it} \mid k_{it}, l_{it}, m_{it}, m_{it-1}, k_{it-1}, u_{it}] = 0, \qquad \mathbb{E}[\xi_{it} + \epsilon_{it} \mid k_{it}, m_{it-1}, k_{it-1}] = 0.$$

In this estimation, the parameters might be identified even if labor is a flexible input, expressed as $l_{it} = l(\omega_{it}, k_{it})$. The main difference between the control variable approach of Imbens and Newey (2009) and the proxy variable approach lies in their conditioning variables: the proxy approach conditions on an unknown function of (k_{it}, m_{it}), whereas the control variable approach conditions on u_{it}, a known function of (k_{it}, m_{it}). Even conditional on u_{it}, there may still be variation in l_{it} linearly independent of k_{it}.

To illustrate this, suppose the firm's labor choice is $l_{it} = l\big(k_{it}, \omega_{it}^H\big) = l\big(k(k_{it-1}, \omega_{it-1}^H, \nu_{it-1}), c(m_{it-1}, k_{it-1}, u_{it})\big)$ where $k_{it} = k(k_{it-1}, \omega_{it-1}, \nu_{it-1})$ defines the next-period capital via the firm's investment decision, and ν_{it-1} is a vector of random variables (e.g., investment prices) influencing investment decision. Substituting $\omega_{it-1} = s^{-1}(k_{it-1}, m_{it-1})$ we obtain $l_{it} = l\big(k(k_{it-1}, s^{-1}(k_{it-1}, m_{it-1}), \nu_{it-1})), c(m_{it-1}, k_{it-1}, u_{it})\big)$, which yields $l_{it} =: \tilde{l}(k_{it-1}, m_{it-1}, u_{it}, \nu_{it-1})$. In this example, even when conditioning on $(k_{it-1}, m_{it-1}, u_{it})$, the variables ν_{it-1} could generate variation in labor independently of a linear function of capital. Note, however, that this illustration depends on the parametric form of the production function.

D Additional Proofs

Proof of Proposition 4.3

In this proof, we omit time subscripts from all functions for notational simplicity. This proof proceeds in two parts. First, I show that there exist two different sets of structural functions, $h(\cdot)$ and $\bar{r}(\cdot)$, that produce the same values for $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. Second, I show that labor-augmenting productivity, the output elasticity of capital, and the elasticity of substitution depend on $h(\cdot)$ and $\bar{r}(\cdot)$, implying that these quantities cannot be uniquely inferred from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. Starting with the elasticities, note that θ_{it}^L and θ_{it}^M depend on the production function as follows:

$$\theta_{it}^L = f_2(\cdot) h_2\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big) \bar{r}(K_{it}, \tilde{M}_{it}) L_{it} \tag{OA.21}$$

$$\theta_{it}^{M} = f_2(\cdot)h_3\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big)M_{it}, \tag{OA.22}$$

where the arguments of $f_2(\cdot)$'s derivatives are omitted. Next, the derivatives of the reduced form function $\bar{h}(\cdot)$ are expressed as follows:

$$\bar{h}_2(K_{it}, \tilde{M}_{it}) = h_2\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big)\bar{r}_2(K_{it}, \tilde{M}_{it}) + h_3\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big), \tag{OA.23}$$

$$\bar{h}_1(K_{it}, \tilde{M}_{it}) = h_1\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big) + h_2\big(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}\big)\bar{r}_1(K_{it}, \tilde{M}_{it}). \tag{OA.24}$$

The left-hand sides of Equations (OA.21-OA.24) are functions of $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. This raises the question of whether the unknown quantities on the right-hand sides of Equations (OA.21-OA.24) can be identified solely from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$. To investigate this, consider the true functions $h(\cdot), \bar{r}(\cdot)$ and alternatives $h^*(\cdot), \bar{r}^*(\cdot)$ defined as:

$$\begin{aligned}
\bar{r}^*(K_{it}, \tilde{M}_{it}) &= \bar{r}(K_{it}, \tilde{M}_{it})T(K_{it}), \\
h_2^*(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})/T(K_{it}), \\
h_1^*(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) - \bar{r}(K_{it}, \tilde{M}_{it})T'(K_{it})/T(K_{it}), \\
h_3^*(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) &= h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}),
\end{aligned}$$

where $T(K_{it})$ is an arbitrary function and $T'(K_{it})$ is its derivative. These true and alternative functions lead to the same $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$:

$$\begin{aligned}
\theta_{it}^L/f_2 &= h_2^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}^*(K_{it}, \tilde{M}_{it})L_{it} \\
&= h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}(K_{it}, \tilde{M}_{it})L_{it}, \\
\theta_{it}^M/f_2 &= h_3^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it} = h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})M_{it}, \\
\bar{h}_2(K_{it}, \tilde{M}_{it}) &= h_2^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}_2^*(K_{it}, \tilde{M}_{it}) + h_3^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}), \\
&= h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}_2(K_{it}, \tilde{M}_{it}) + h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}), \\
\bar{h}_1(K_{it}, \tilde{M}_{it}) &= h_1^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2^*(K_{it}, \bar{r}^*(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}_1^*(K_{it}, \tilde{M}_{it}), \\
&= h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) + h_2(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})\bar{r}_1(K_{it}, \tilde{M}_{it}).
\end{aligned}$$

Hence, using $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$ alone does not allow us to distinguish the original functions $(h_1, h_2, \bar{r}_1, \bar{r}_2)$ from the alternatives $(h_1^*, h_2^*, \bar{r}_1^*, \bar{r}_2^*)$. Next, I will show that labor-augmenting productivity, capital elasticity, and the elasticity of substitution all depend on $(h_1, h_2, \bar{r}_1, \bar{r}_2)$, so they cannot be inferred from $(\theta_{it}^L, \theta_{it}^M, \bar{h}, f)$.

Since $\omega_{it}^L = \bar{r}(K_{it}, \tilde{M}_{it})$, non-identification of $\bar{r}(K_{it}, \tilde{M}_{it})$ immediately implies that ω_{it}^L is not identified. The capital elasticity depends on $h_1(\cdot)$, $\theta_{it}^K =$

$f_1 + f_2 h_1(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$, and therefore is not identified. Finally, to demonstrate that the elasticity of substitution is not identified, observe that it is defined as $\sigma_{it}^{ML} = \partial \log(L_{it}/M_{it})/\partial \log(F_M/F_L)$ where the ratio of marginal products is computed as:

$$\frac{F_L}{F_M} = \frac{h(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})}{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})} - \tilde{M}_{it}.$$

Using this, the elasticity of substitution between materials and labor is given by

$$\sigma_{it}^{ML} = \frac{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})^2 - h(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it}) h_{33}(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})}{h_3(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})^2} - 1,$$

which depends on $h_{33}(K_{it}, \bar{r}(K_{it}, \tilde{M}_{it}), \tilde{M}_{it})$. This function is not identified because $\bar{r}(K_{it}, \tilde{M}_{it})$ and $h(\cdot)$ are not identified. Therefore, σ_{it}^{ML} is not identified. The elasticity of substitution for other input pairs can similarly be derived, and it can be shown that they are not identified because they depend on the derivatives of the function $h(\cdot)$.

Proof of Proposition 4.4

I drop time subscripts from all functions for brevity. If the production function takes the form given in Equation (4.7), the labor and materials elasticities, as a function of f and h, can be written as

$$\theta_{it}^L = f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}(\tilde{M}_{it}) L_{it}, \qquad \theta_{it}^M = f_2 h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) M_{it}.$$

Since I already showed in Equation (4.6) that θ_{it}^L and θ_{it}^M are identified, the right-hand sides of these equations are identified. The identification of θ_{it}^M immediately implies that $h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$ is identified from (f_2, θ_{it}^M). Taking the derivative of the reduced form function $\bar{h}(\cdot)$ with respect to $\tilde{M}_{it}$, and using $\bar{h}(\tilde{M}_{it}) = h(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})$, I obtain

$$\bar{h}'(\tilde{M}_{it}) = h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it}) + h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}), \tag{OA.25}$$

where $\bar{r}'(\tilde{M}_{it})$ denotes the derivative of $\bar{r}(\tilde{M}_{it})$. Therefore, the right-hand side of Equation (OA.25) is identified from $\bar{h}(\tilde{M}_{it})$ by taking its derivative. Taking the ratio of $f_2 \bar{h}'(\tilde{M}_{it}) - \theta_{it}^M / M_{it}$ and θ_{it}^L / L_{it}, and denoting it by $b(\tilde{M}_{it})$:

$$b(\tilde{M}_{it}) := \frac{f_2 \bar{h}'(\tilde{M}_{it}) - \theta_{it}^M / M_{it}}{\theta_{it}^L / L_{it}} = \frac{f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}'(\tilde{M}_{it})}{f_2 h_1(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it}) \bar{r}(\tilde{M}_{it})} = \frac{\bar{r}'(\tilde{M}_{it})}{\bar{r}(\tilde{M}_{it})} = \frac{\partial \log(\bar{r}(\tilde{M}_{it}))}{\partial \tilde{M}_{it}}.$$

which implies that the derivative of $\log(\bar{r}(\tilde{M}_{it}))$ with respect to $\tilde{M}_{it}$ is identified from $(\theta^L_{it}, \theta^M_{it}, \bar{h}, f)$. Thus, by integrating $b(\tilde{M}_{it})$, we can recover $\log(\bar{r}(\tilde{M}_{it}))$ up to an additive constant:

$$\log(\bar{r}(\tilde{M}_{it})) = \int_{\underline{\tilde{M}}}^{\tilde{M}_{it}} b(\bar{M})d\bar{M} + a.$$

Since $\omega^L_{it} = \bar{r}(\tilde{M}_{it})$, and $\log(\bar{r}(\tilde{M}_{it}))$ is identified up to a constant, ω^L_{it} is identified up to a scale. Capital elasticity is also identified because it depends only on the reduced form functions f and $\bar{h}$: $\theta^K_{it} = f_1(K_{it}, L_{it}\bar{h}(\tilde{M}_{it}))K_{it}$.

Proof of Proposition 4.5

If the production function takes the form in Equation (4.7), then we can derive σ^{ML}_{it} as

$$\sigma^{ML}_{it} = \frac{h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})^2 - h(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})h_2(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})}{h_{22}(\bar{r}(\tilde{M}_{it}), \tilde{M}_{it})^2} - 1,$$

which depends on $h_{22}(\cdot)$. Since $h_{22}(\cdot)$ is not identified, σ^{ML}_{it} is not identified. Similarly, one can derive the elasticities of substitution for other input pairs, which also depend on the second derivatives of $h(\cdot)$.

D.1 Identification of the Reduced Form Functions Under Strongly Homothetic Production Function

In this section, I analyze the identification of the reduced form functions $(f_t, \bar{h}_t)$ obtained from the weakly homothetic production function given in Equation (2.2). This analysis follows the proof of Proposition 5.1 closely. While Proposition 5.1 showed identification of reduced form functions for the strongly homothetic production function $F_t\big(K_{it}, h_t(\omega^L_{it}L_{it}, M_{it})\big)$ this section analyzes identification for the weakly homothetic production function $F_t\big(K_{it}, h_t(K_{it}, \omega^L_{it}L_{it}, M_{it})\big)$.

Since Section 4.2 showed that under the weakly homothetic production function, only the sum of labor and materials elasticities is identifiable from the reduced form production functions as $\theta^V_{it} := f_{t2}\big(K_{it}, L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it})\big)L_{it}\bar{h}_t(K_{it}, \tilde{M}_{it})$, I focus on identification of $f_{t2}\bar{h}_t$ using the moment condition given in Equation (5.3). This analysis follows the proof of Proposition 5.1 closely. Under this functional form, Equation (B.10) of Proposition 5.1 becomes

$$\mathbb{E}[y \mid w, z, x, t] = f\big(w, zh(w, x)\big) + g\big(t, l(w, x, t), s(w, z, x, t)\big), \tag{OA.26}$$

where $l(w,x,t) = F_{\tilde{M}_{it}|K_{it},W_{it-1}}(\tilde{M}_{it} \mid K_{it}, W_{it-1})$ as derived in Equation (3.3) and $(w,z,x,t) := (K_{it}, L_{it}, \tilde{M}_{it}, W_{it-1})$. $s(\cdot)$ and $g(\cdot)$ are defined in Proposition 5.1. Taking the derivative of Equation (OA.26) with respect to w, z and x, we obtain

$$y_1(w,z,x,t) = f_1\big(w, zh(w,x)\big)h_1(w,x)z + g_3 s_1(w,z,x,t) + g_2 l_1(w,x,t) \quad \text{(OA.27)}$$

$$y_2(w,z,x,t) = f_2\big(w, zh(w,x)\big)h(w,x) + g_3 s_2(w,z,x,t) + g_2 l_2(w,x,t), \quad \text{(OA.28)}$$

$$y_3(w,z,x,t) = f_2\big(w, zh(w,x)\big)h_2(w,x)z + g_3 s_3(w,z,x,t) \quad \text{(OA.29)}$$

where the arguments of g functions are suppressed for brevity. Summing these equations after multiplying Equation (OA.27) by $\alpha = -l_2/l_1$ and multiplying Equation (OA.29) by $\beta = (l_2/l_1)s_1/s_3 - s_2/s_3$, we obtain

$$\alpha y_1 + y_2 + \beta y_3 = \alpha f_1(w, zh(w,x))h_1(w,x)z + f_2(w, zh(w,x))h(w,x) + \beta f_2(w, zh(w,x))h_2(w,x)z$$

which can be written in the form of

$$\widehat{y}(w,z,x,t) = (\alpha+\beta)\widehat{f}(w,z,x) + \widehat{f_1}(w,z,x), \quad \text{(OA.30)}$$

where $\widehat{f}(w,z,x) := f_1(w, zh(w,x))h_1(w,x)z + f_2(w, zh(w,x))h_2(w,x)z$ and $\widehat{f_1}(w,z,x) := f_2(w, zh(w,x))h(w,x)$. Here $\widehat{y}$, α and β are known functions and $\widehat{f}$ and $\widehat{f_1}$ are unknown functions. Lemma B.1 implies that $\widehat{f_1}(w,z,x)$ is identified from this Equation (OA.30) under the assumption that $\mathrm{Var}\big(\alpha+\beta \mid w,z,x\big) > 0$. By Proposition 4.2, the sum of flexible input elasticities equals $\widehat{f_1}(w,z,x)z = f_2(w, zh(w,x))h(w,x)z$. Thus, the sum of elasticities is identified.

E Estimation Details

E.1 Estimation Algorithm

This section outlines the estimation algorithm. First, we apply the data cleaning and variable construction steps described in Online Appendix F and denote the resulting sample as A. Next, remove any observations with missing lagged inputs to form sample B. From sample B, extract the subset of observations that lies within the rolling window τ, denoting this new subsample B_τ. To estimate the control variable u^2_{it} for each $it \in B_\tau$, partition the support of M_{it} into 500 equally-sized bins (by observation count). Let Q denote the set of these grid points. For each $q \in Q$, use logistic regression with a third-degree polynomial in (k, w, u) to estimate

$$\Pr(M_{it} \leqslant q \mid K_{it} = k, W_{it-1} = w, u_{it}^1 = u) \equiv s_t(q, k, w, u)$$

Then, for each $it \in B_\tau$, compute an estimate $\widehat{u}_{it}^2$ of $u_{it}^2 = s_t(M_{it}, K_{it}, W_{it-1}, u_{it}^1)$ by linearly interpolating between the two grid points in Q nearest to M_{it}. This procedure yields estimates $\widehat{u}_{it}^2$ for all observations in B_τ. To estimate the production function, first approximate $\log(\bar{h}_t)$ using a third-degree polynomial:

$$\log(\widehat{\bar{h}}_t(\tilde{M}_{it})) = a_{1t} + a_{2t}\tilde{m}_{it} + a_{3t}\tilde{m}_{it}^2 + a_{4t}\tilde{m}_{it}^3, \tag{OA.31}$$

where $\{a_{jt}\}_{j=1}^4$ are the parameters of the polynomial approximation. Set $a_{1t} = 0$ to impose a normalization because $\bar{h}_t$ is identified up to a multiplicative constant. Define $V_{it} := L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})$ and v_{it} its logarithm. Approximate the production function as follows:

$$\begin{aligned}\widehat{f}_t(K_{it}, L_{it}\widehat{\bar{h}}_t(\tilde{M}_{it})) = b_{1t} + b_{2t}k_{it} + b_{3t}k_{it}^2 + b_{4t}k_{it}^3 + b_{5t}v_{it} + b_{6t}v_{it}^2 + b_{7t}v_{it}^3 \\ + b_{8t}k_{it}^2 v_{it} + b_{9t}k_{it}v_{it}^2 + b_{10t}k_{it}v_{it},\end{aligned} \tag{OA.32}$$

where $\{b_{jt}\}_{j=1}^{10}$ are the polynomial coefficients. Approximate the control functions $c_{2t}(\cdot)$ and $c_{3t}(\cdot)$ using a similar third-degree polynomial approach. For given values of $\{a_{jt}\}_{j=1}^4$, $\{b_{jt}\}_{j=1}^{10}$, $\widehat{c}_{2t}(\cdot)$ and $\widehat{c}_{3t}(\cdot)$, construct the objective function from Equation (5.5) using the third-degree polynomials of K_{it} and W_{it-1} as $\{z_j(K_{it}, W_{it-1})\}_{j=1}^J$ instruments. Estimate the production function by minimizing this objective function using the following nested optimization procedure. In the inner loop, for each value of $\{a_{jt}\}_{j=2}^4$, use least squares regression to estimate $\{b_{jt}\}_{j=1}^{10}$, $\widehat{c}_{2t}(\cdot)$ and $\widehat{c}_{3t}(\cdot)$. In the outer loop, apply an optimization routine to estimate $\{a_{jt}\}_{j=1}^4$. After estimating the production function parameters, the next step is elasticity and markup estimation.

Select observations at the midpoint of the rolling window in sample A to form A_c. For each $it \in A_c$, calculate output elasticities and markups as follows: first obtain estimates of $f_t(\cdot)$ and $\bar{h}_t(\cdot)$ from the parameter estimates $\{a_{jt}\}_{j=2}^4$ and $\{b_{jt}\}_{j=1}^{10}$ in Equations (OA.31) and (OA.32). With the estimates of $f_t(\cdot)$ and $\bar{h}_t(\cdot)$, compute the capital elasticity of output and the sum of the materials and labor elasticities, as specified in Equations (4.5) and (4.9). Next, given the estimated $\widehat{\theta}_{it}^V$ and the revenue shares for labor and materials, use Equation (4.6) to estimate the output elasticities of labor and materials. Then, calculate markups using $\widehat{\theta}_{it}^V$

and the revenue share of flexible input. To compute standard errors, resample firms with replacement from sample A 250 times and repeat the entire estimation process.

I follow the same procedure to estimate the CES and nested CES models, applying the necessary parametric restrictions.

E.2 Blundell and Bond (2000) Estimation Method

In this section, I present the dynamic panel method of Blundell and Bond (2000), which I use in Section 8 to estimate Hicks-neutral production functions. Under the dynamic panel approach, productivity shocks follow an AR(1) process: $\omega_{it}^H = \rho\omega_{it-1}^H + \upsilon_{it}$. Given this assumption, we take the ρ-difference of the log production function, which yields:

$$\begin{aligned} y_{it} - \rho y_{it-1} =& \beta_k(k_{it} - \rho k_{it-1}) + \beta_l(l_{it} - \rho l_{it-1}) + \beta_m(m_{it} - \rho m_{it-1}) + \\ & \omega_{it}^H - \rho\omega_{it-1}^H + \epsilon_{it} - \rho\epsilon_{it-1}. \end{aligned}$$

Let $\nu_{it} := \omega_{it}^H - \rho\omega_{it-1}^H + \epsilon_{it} - \rho\epsilon_{it-1}$ be the composite error term. By construction, ν_{it} is orthogonal to the firm's information set at $t-1$, that is, $\mathbb{E}[\nu_{it} \mid \mathcal{I}_{it-1}] = 0$. To exploit this orthogonality condition, define the following moment function:

$$\nu_{it}(\beta_k, \beta_l, \beta_m, \rho) = y_{it} - \rho y_{it-1} - \beta_k(k_{it} - \rho k_{it-1}) - \beta_l(l_{it} - \rho l_{it-1}) - \beta_m(m_{it} - \rho m_{it-1}).$$

The moment conditions to estimate the parameters are given by:

$$\mathbb{E}\Big[\nu_{it}(\beta_k, \beta_l, \beta_m, \rho)(k_{it}, k_{it-1}, l_{it-1}, m_{it-1})'\Big] = 0.$$

I estimate the translog production function similarly, imposing the following functional form:

$$\begin{aligned} y_{it} = & \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + \beta_4 k_{it}^2 + \beta_5 l_{it}^2 + \beta_6 m_{it}^2 + \\ & \beta_7 k_{it} m_{it} + \beta_8 l_{it} m_{it} + \beta_9 l_{it} k_{it} + \omega_{it}^H + \epsilon_{it}. \end{aligned}$$

F Data Appendix

This section describes the datasets. The summary statistics for each dataset are presented in Supplementary Appendix A.

F.1 Chile

Data from Chile are from the Chilean Annual Census of Manufacturing, Encuesta Nacional Industrial Anual (ENIA), covering the years 1979 through 1996. This dataset includes all manufacturing plants with at least ten employees. I restrict the sample to industries with more than 250 firms per year. I drop observations at the bottom and top 2% of the distribution of revenue share of labor, revenue share of materials, or the combined flexible input for each industry to remove outliers. I report each industry's share in manufacturing in terms of sales and the number of plants for the first, last, and midpoint year of the sample in Table SA-1. After sample restrictions, five industries remain, covering around 30% of the manufacturing sector.

F.2 Colombia

The data for Colombia are from the annual Colombian Manufacturing census provided by the Departamento Administrativo Nacional de Estadistica, covering the years 1981 through 1991. This dataset contains all manufacturing plants with at least ten employees. I restrict the sample to industries with more than 250 firms per year on average and follow the same steps as those used for constructing the Chile data to remove outliers. Table SA-2 shows that the number of industries in the sample is nine, and the sample covers around 55% of the manufacturing sector's total sales.

F.3 India

The Indian data were collected by the Ministry of Statistics and Programme Implementation through the Annual Survey of Industries (ASI), which covers all plants with at least ten workers and those that use electricity, as well as those that do not use electricity but have at least 20 workers. The plants are divided into two categories: census and sample. The census sector comprises all large plants and all plants in states classified as industrially backward by the government. From 2001 to 2005, a large plant was defined as one with 200 or more workers; however, starting in 2006, the definition was revised to one with 100 or more workers. All plants in the census sector are surveyed every year. The remaining plants constitute the sample sector, from which a random sample is surveyed each year. India uses the National Industrial Classification (NIC) to categorize manufacturing establishments, a classification system similar to those used in other countries. The industry definition has changed multiple times over the sample period. I follow Allcott et al. (2016) to create a consistent industry definition at the NIC 87 level. For sample restrictions and data cleaning, I follow the approach outlined in Allcott et al. (2016). Then, I restrict the sample to Census firms to follow them over time. My final sample comprises industries with an average of more than 250 firms per year. I follow the same steps as in the Chilean data to remove outliers. Table SA-3 provides summary statistics.

F.4 Turkey

Manufacturing data for Turkey come from the Turkish Statistical Institute (TurkStat), which maintains comprehensive plant-level records for the sector. TurkStat conducts the Annual Surveys of Manufacturing Industries (ASMI), covering establishments with ten or more employees. I use a sample of ASMI for the period from 1983 to 2000. The main variables include gross revenue, investment, the book value of capital, materials expenditures, and the number of production and administrative workers. For variable construction, I follow Taymaz and Yilmaz (2015). I restrict the sample to industries with more than 250 firms per year on average and private establishments. I follow the same procedure as I do with Chilean data to remove outliers.

F.5 Compustat

The Compustat data comes from Standard and Poor's Compustat North America database, covering a period from 1961-2018. Although later data are available, the sample ends in 2018 because I use the NBER-CES Manufacturing Industry Database to deflate values in Compustat, which is available only until 2018.[45] Since Compustat draws from corporate financial statements, it requires more extensive data cleaning compared to other datasets. The cleaning process involves removing non-US incorporated firms, financial and utility companies (industry codes 4900-4999 and 6000-6999), firms with negative or zero values for key variables (sales,

[45] The NBER-CES Manufacturing Industry Database can be found at https://www.nber.org/research/data/nber-ces-manufacturing-industry-database.

employment, cost of goods sold, selling/general/administrative expenses), companies with fewer than ten employees, and those without industry codes. The final sample includes only manufacturing firms with NAICS codes 31-33. Variable construction follows the methodology detailed in Keller and Yeaple (2009), specifically their Appendix B, page 831. Unlike the plant-level datasets, Compustat provides firm-level data and covers only public companies. Industries are classified using two-digit NAICS codes. Outlier removal excludes the top and bottom 1% of observations, rather than the 2% threshold used in other datasets, to preserve the sample size.

F.6 Variable Construction

Labor: For Chile, Colombia, Turkey, and the United States, labor is measured using the number of production workers, while for India, it is measured using the total number of days worked. The labor revenue share is calculated as the sum of total salaries and benefits divided by total sales for the year.

Materials: For Chile, Colombia, India, and Turkey, materials input is computed as the sum of total materials expenditures and the year-over-year change in inventory value. This nominal value is converted to real terms using the industry-level intermediate input price index. For Compustat, materials are calculated as the deflated sum of cost of goods sold and administrative/selling expenses, minus depreciation and wage expenditures. The materials revenue share represents materials cost divided by total annual sales.

Capital: For Turkey, the capital stock is constructed using the perpetual inventory method, combining new investment with deflated capital from the last period to determine current capital levels. For Compustat, capital is measured as the deflated value of property, plant, and equipment net of depreciation, using BEA satellite account deflators. For India, the capital book value is deflated using an implied national deflator derived from "Table 13: Sector-wise Gross Capital Formation" in the Reserve Bank of India's Handbook of Statistics on the Indian Economy. For Chile and Colombia, I follow Raval (2023).

Output: The output measure is calculated as deflated sales. For Compustat, it is net sales from the Industrial data file. For other countries, sales are the total production value plus the difference between the end-year and beginning-year value of finished goods inventories.

G Robustness Checks Details

G.1 Estimation with Physical Quantity

One challenge in estimating production functions with physical quantities is the aggregation of products for multi-product firms. To resolve this issue, I focus on single-product firms and estimate the production functions of six relatively homogeneous products in Indian manufacturing industries. For this sample, I observe both the quantity of production and the price. These products include Brick Tiles, Cotton Shirts, Biri Cigarettes, Black Tea, Parboiled Non-Basmati Rice, and Raw Non-Basmati Rice. In this sample, products are relatively homogeneous and produced by a sufficient number of plants. Table SA-6 lists the products, their units, the number of firm-year observations, and their product codes. The product classification includes two types of codes, ASICC and NPCMS, as the National Statistical Office of India revised its product coding system in 2011.[46]

To construct the sample, I follow Raval (2023) and include only plants that derive at least 75% of their revenues from one of these products. Product price for each firm is calculated by subtracting reported sales-related expenses (excise duty, sales tax, and other expenses) from gross value and dividing it by the quantity sold. To remove outliers, I exclude plants whose prices are greater than five times or less than 20% of the median price for a given product-year. Because each industry in this sample has fewer observations than in the main sample, I use a five-year rolling window instead of a three-year window.

With this sample, I estimate product-level production functions. The main difference from the estimation in the main text is that I use physical output quantities instead of deflated revenues for Y_{it}. All other estimation procedures follow the methodology outlined in Online Appendix E.1.

Figure OA-7a presents the comparison of elasticities and markups when estimated with revenues and quantities for the FA production function (left) and markup estimates from different production functions (right). Elasticity comparisons suggest that using deflated revenues instead of quantity introduces an upward

[46] The product codes after 2010 can be found here. The crosswalk between the two product categories can be found here.

bias in capital elasticity (7.2%), a downward bias in variable elasticity (-6.5%), and a negligible bias in markups (0.1%). Markup comparison in the right panel reveals the same pattern documented in the main text: Hicks-neutral production functions yield the highest markup estimates, and markup estimates decline progressively as we move from Cobb-Douglas to nonparametric labor-augmenting productivity specifications.

G.2 Estimation when Controlling for Input Prices

This robustness check includes input prices for materials and wages as controls in the input demand functions as described in the extension in Online Appendix A.

The Indian manufacturing data provide the nine largest inputs for each plant, along with their product code, quantity, value, price, and unit. An "other" category encompasses inputs beyond these nine. For each of the six industries used in the quantity production function estimation in Online Appendix G.1, I identify the largest input by expenditure share. These main inputs remain consistent across the years; however, their codes and, sometimes, their names changed in 2010 due to changes in the product classification coding system. To address this, I match the ASICC (2001-2009) and NPCMS (2010-2014) codes for these inputs using the ministry's concordance. I exclude the year 2000 from the analysis because the largest input is listed as "Other basic items (Indigenous)" for almost all industries.

I identify the input unit most commonly used across firms, as shown in Table SA-7. Most firms consistently use the same unit over time. For firms using different units, I perform appropriate conversions (e.g., kg to tonnes), which yield the input set and corresponding prices. For a small fraction of firms (5-15% depending on industry) where the price of the main input is unavailable, I substitute the industry-year average price.

Table SA-7 shows summary statistics for the inputs considered in the paper. The largest input typically accounts for a large fraction of intermediate input expenditures, from 53% for shirts manufacturing to 90% for non-basmati rice. The input price variation is relatively small, as indicated by the average CoV in each industry over the years, except for cigarettes and brick tiles.

I also calculate the average wage by dividing the total wage bill by the total number of employees. With these materials prices and wages, I follow the estimation procedure in Online Appendix A and estimate markups by controlling for variation in input prices.

Results in Figure OA-7b suggest that capital elasticity is biased downward (-2.3%) and variable elasticity is biased downward when using revenues instead of quantities. There is also a -0.9% bias in average markup estimates. However, the markup comparison results across different production functions are consistent with the pattern documented in the main text.

G.3 Measurement Error in Capital

This section employs a simulation study to examine the impact of potential measurement errors in capital input on empirical estimates. I assume that the observed data come from the 'true' data-generating process and introduce errors to the observed data to generate simulated data with measurement error in capital. Specifically, I add an independent and identically distributed error term to the capital input, drawn from a mean-zero normal distribution with a standard deviation of one-tenth of the capital's standard deviation in the data. After simulating 100 datasets with these measurement errors, I estimate output elasticities and markups for each dataset and report their averages across 100 simulations.

Figure OA-9 presents results alongside the original estimates. I find that measurement error in capital reduces the output elasticity of capital and increases the output elasticity of labor in most simulations. This suggests that the higher capital elasticity estimates obtained using the factor-augmenting production function in Section 6.1 are unlikely to be driven by the impact of measurement error in capital.

To examine how potential capital underutilization affects the empirical estimates, I analyze firms' energy consumption, assuming a Leontief production function relationship between capital and energy. This assumption enables me to infer the actual capital utilized by firms through their energy consumption, as these inputs are expected to be proportionally related in a Leontief production function. Since energy consumption data are available in the Chilean and Turkish datasets, I restrict this robustness analysis to firms from these two countries. I first calculate the true capital used by each firm using the electricity data, and then use these corrected capital values instead of the observed values in production function estimation.

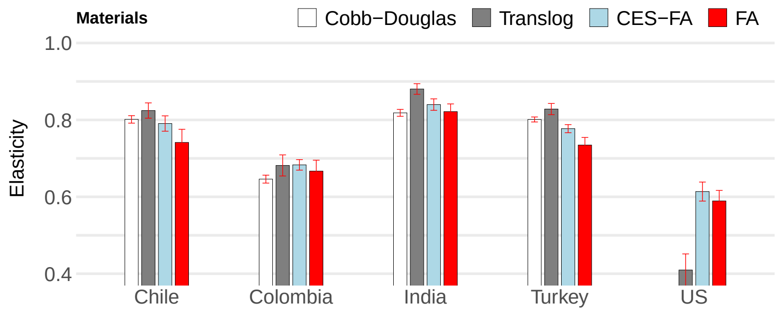
Figure OA-8 compares the original estimates with those obtained using capacity utilization-adjusted capital. The results indicate that this adjustment primarily affects capital elasticities, while other elasticities and markups show small differences in magnitude but maintain similar patterns across estimates. In Turkey, the capacity utilization correction preserves the comparative patterns in capital elasticity estimates across production function specifications. However, in Chile, the adjusted capital elasticity estimates are too small, possibly due to noise in the electricity consumption data.

G.4 Estimation with Single Product Firms

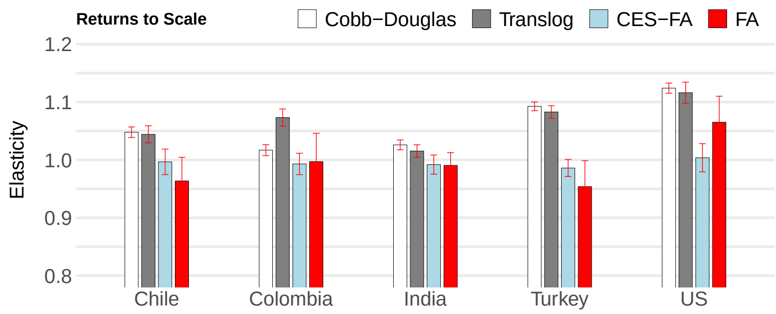
When I estimate the quantity production function, I include plants that obtain at least 75% of their revenue from a given product. This procedure still includes firms that are multi-product, which can bias the markup and production function estimates. As a robustness check, I include the plants that receive at least 99% of their revenue from a given product, focusing on purely single-product firms. The results reported in Figure OA-7c suggest that capital elasticity is biased downward (-10.7%) and variable elasticity is biased downward (-6.0%) when using revenues instead of quantities. There is also a 4.5% bias in average markup estimates. However, the markup comparison results across different production functions are consistent with the pattern documented in the main text.

H Additional Empirical Results

Figure OA-1: Additional Output Elasticity Estimates



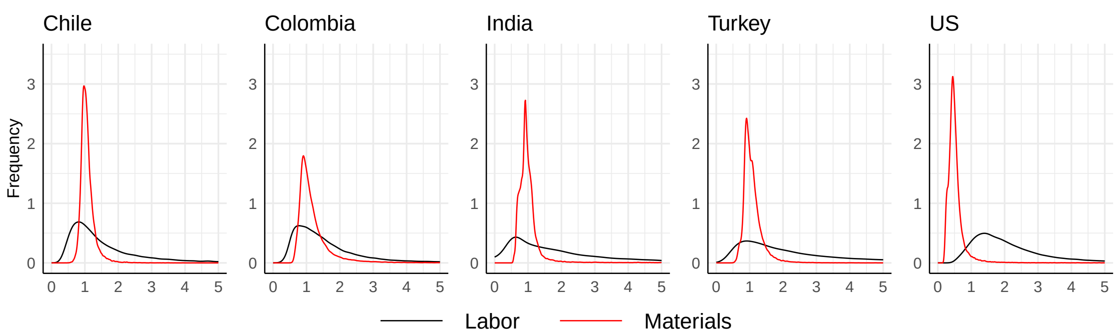
(a) Materials Elasticity



(b) Returns to Scale

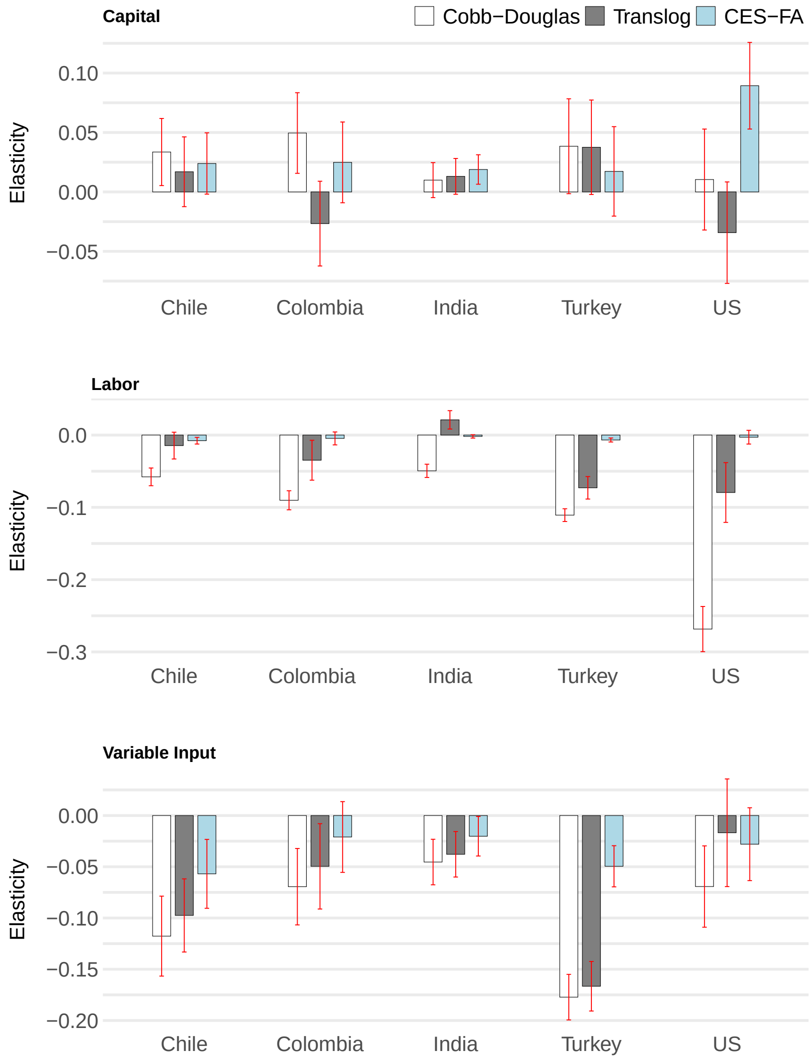
Notes: Comparison of sales-weighted materials elasticities and returns to scale in each country generated by the following production functions: (i) Cobb-Douglas (CD), (ii) Translog (TR), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) strongly homothetic production function with factor-augmenting productivity (FA). For each year and industry, sales-weighted averages are calculated, and then simple unweighted averages are taken over the years. The error bars indicate 95 percent confidence intervals calculated using bootstrap method (250 resamples).

Figure OA-2: Distribution of Markups from Labor and Materials (Cobb-Douglas)



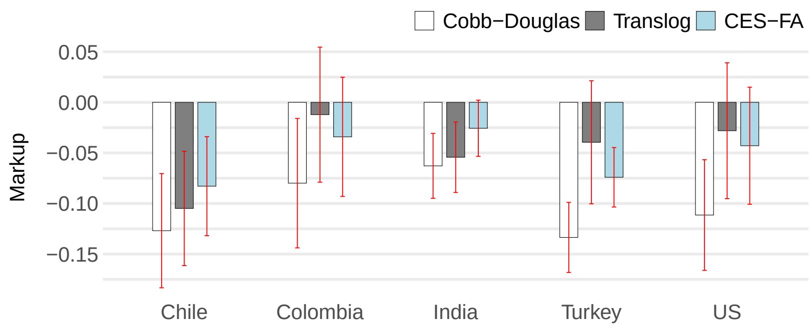
This figure compares the distribution of markups implied by labor (black) and materials (red) elasticities for each country from the Cobb-Douglas specification estimated using the Blundell and Bond (2000) method.

Figure OA-3: Difference of Elasticity Estimates from FA and Other Models



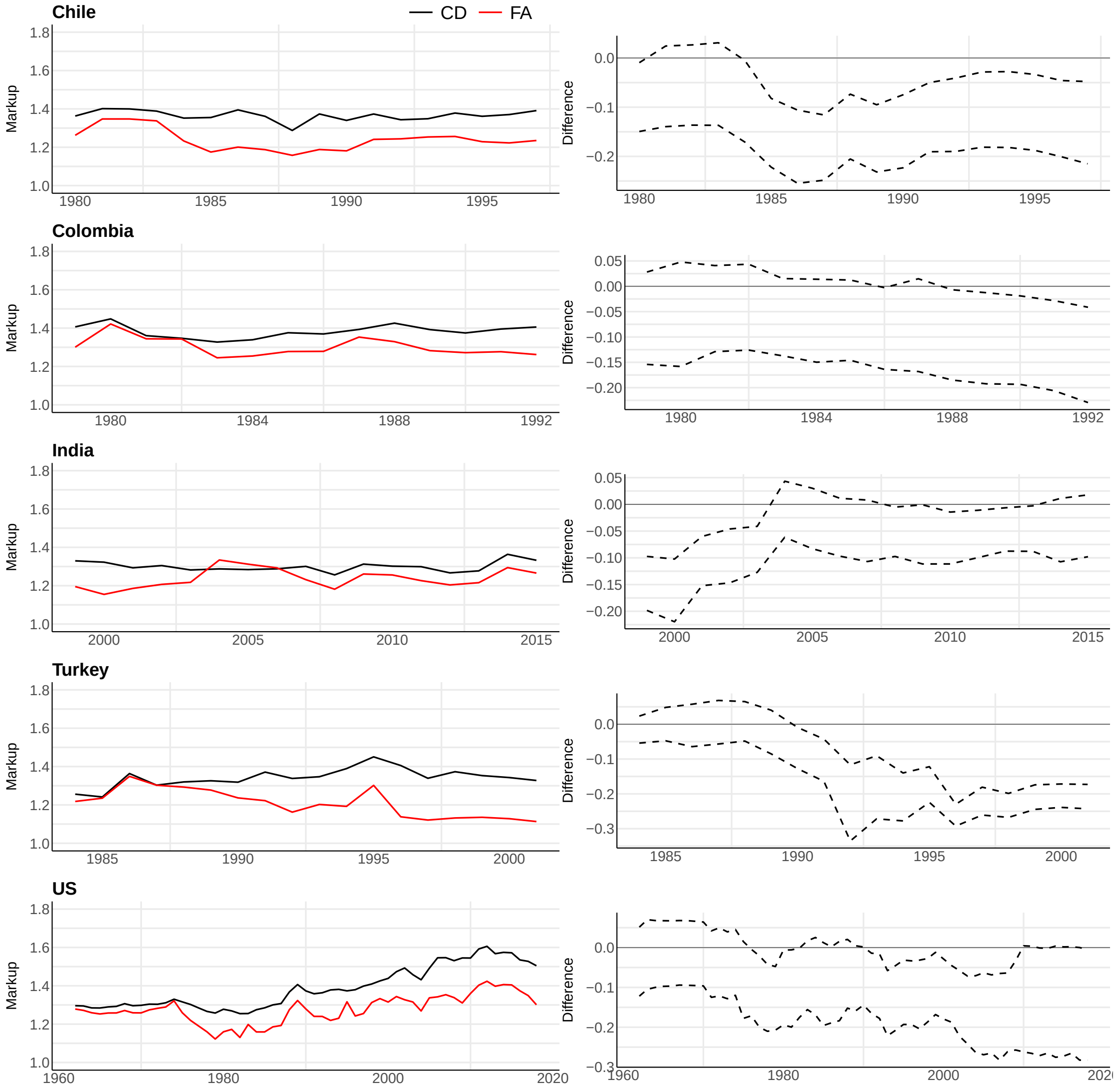
Notes: Each bar shows the difference between the elasticity estimates from the method indicated by the legend color and the strongly homothetic production function with factor-augmenting productivity (FA) method. The red bars represent the corresponding 95 percent confidence intervals, with standard errors computed using bootstrap (250 resamples).

Figure OA-4: Difference of Markup Estimates from FA and Other Models



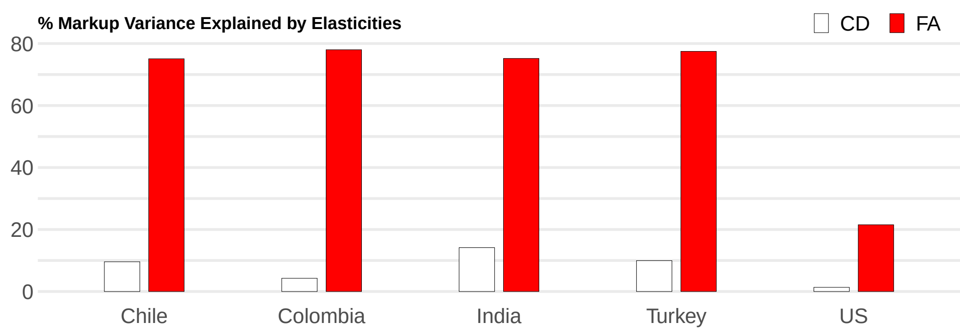
Notes: Each bar shows the difference between the markup estimates from the method indicated by the legend color and the strongly homothetic production function with factor-augmenting productivity (FA) method. The red bars represent the corresponding 95 percent confidence intervals, with standard errors computed using bootstrap (250 resamples).

Figure OA-5: Difference between Markup Estimates and Confidence Bands



Notes: This figure shows changes in aggregate markups estimated using FA and CD (left panels) and the 95 percent confidence intervals for their differences (right panels), calculated from bootstrap distributions (250 resamples).

Figure OA-6: Variance Decomposition of the Aggregate Log Markup Changes



Notes: This figure shows the percent variance explained by elasticity changes in the variance decomposition of aggregate log markup into revenue shares and elasticities, based on $\tilde{\mu}_t = \sum w_{it}\log(\theta_{it}) - \sum w_{it}\log(\alpha_{it})$. The covariance term is excluded, so the results represent the variance of the first term divided by the sum of the variances of the first and second terms.

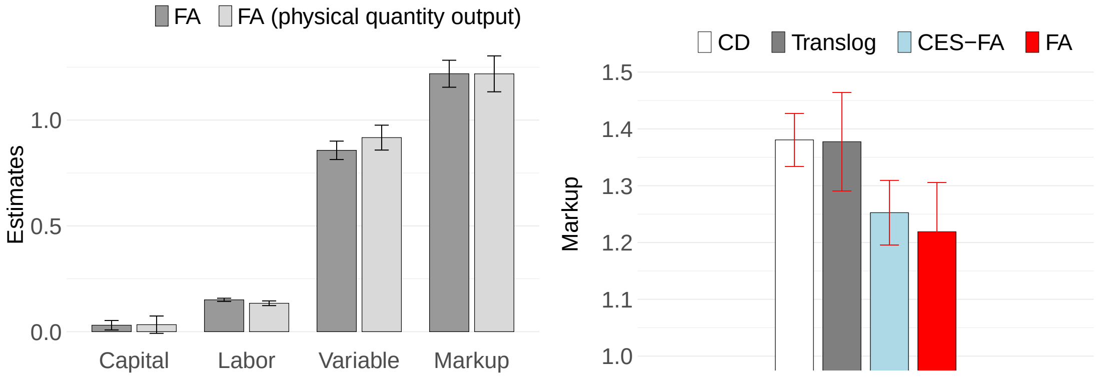
Table OA-1: Output Elasticities in Three Largest Industries in Each Country

	Industry 1				**Industry 2**				**Industry 3**			
	CD	TR	CES	FA	CD	TR	CES	FA	CD	TR	CES	FA
						Chile (311, 381, 321)						
Capital	0.04	0.08	0.06	0.08	0.10	0.07	0.09	0.07	0.11	0.04	0.10	0.10
	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.05)	(0.01)	(0.03)	(0.02)	(0.04)
Labor	0.14	0.09	0.10	0.10	0.23	0.25	0.20	0.18	0.32	0.25	0.19	0.18
	(0.01)	(0.01)	(0.00)	(0.00)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)
Materials	0.86	0.88	0.83	0.78	0.72	0.72	0.70	0.64	0.70	0.75	0.72	0.69
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)
RTS	1.04	1.05	0.99	0.96	1.04	1.03	0.98	0.89	1.13	1.03	1.01	0.97
	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.03)	(0.05)	(0.02)	(0.02)	(0.03)	(0.05)
						Colombia (311, 322, 381)						
Capital	0.06	0.10	0.08	0.14	0.13	0.10	0.07	0.07	0.08	0.30	0.12	0.15
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.05)
Labor	0.18	0.14	0.11	0.12	0.47	0.30	0.31	0.31	0.34	0.30	0.26	0.25
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
Materials	0.79	0.81	0.81	0.83	0.56	0.72	0.67	0.65	0.54	0.52	0.63	0.57
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)
RTS	1.03	1.05	1.01	1.09	1.16	1.12	1.06	1.02	0.96	1.12	1.01	0.97
	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.05)
						India (230, 265, 213)						
Capital	0.09	0.07	0.04	0.03	0.06	0.06	0.05	0.06	0.07	0.05	0.07	0.12
	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Labor	0.34	0.01	0.06	0.06	0.09	0.04	0.08	0.08	0.36	0.14	0.19	0.18
	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.00)
Materials	0.57	0.93	0.85	0.86	0.85	0.91	0.87	0.87	0.57	0.79	0.71	0.68
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
RTS	1.00	1.01	0.95	0.95	1.00	1.01	1.00	1.01	1.00	0.99	0.97	0.97
	(0.02)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
						Turkey (321, 311, 322)						
Capital	0.03	0.05	0.05	0.07	0.05	0.06	0.09	0.10	0.04	0.03	0.05	0.06
	(0.00)	(0.01)	(0.01)	(0.03)	(0.00)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)
Labor	0.16	0.15	0.08	0.07	0.22	0.20	0.15	0.14	0.27	0.18	0.11	0.11
	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
Materials	0.83	0.86	0.82	0.80	0.81	0.81	0.75	0.71	0.71	0.86	0.87	0.86
	(0.00)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
RTS	1.02	1.06	0.94	0.95	1.09	1.07	0.99	0.95	1.02	1.06	1.03	1.03
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.02)	(0.04)
						US (33, 32, 31)						
Capital	0.28	0.46	0.26	0.27	0.23	0.20	0.15	0.22	0.21	0.29	0.12	0.25
	(0.03)	(0.06)	(0.05)	(0.05)	(0.02)	(0.03)	(0.02)	(0.04)	(0.01)	(0.02)	(0.02)	(0.03)
Labor	0.50	0.28	0.20	0.21	0.47	0.33	0.21	0.21	0.54	0.31	0.28	0.27
	(0.04)	(0.06)	(0.02)	(0.01)	(0.03)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Materials	0.25	0.30	0.55	0.57	0.31	0.48	0.64	0.60	0.26	0.37	0.61	0.59
	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
RTS	1.02	1.04	1.01	1.05	1.01	1.02	1.00	1.03	1.01	0.97	1.01	1.11
	(0.01)	(0.02)	(0.03)	(0.05)	(0.01)	(0.02)	(0.02)	(0.04)	(0.00)	(0.01)	(0.01)	(0.02)

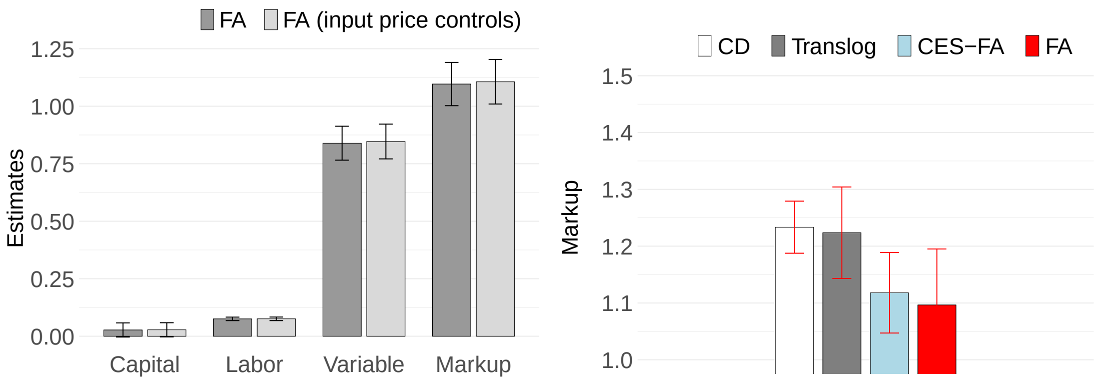
Notes: This table presents a comparison of sales-weighted average output elasticities estimated using different methods: (i) Cobb-Douglas (CD), (ii) Translog (TR), (iii) CES with labor-augmenting productivity (CES-FA), and (iv) strongly homothetic production function with factor-augmenting productivity (FA). For each year and industry, sales-weighted averages are first calculated, followed by taking simple averages across years. Numbers in each panel correspond to the SIC codes of the three largest industries in each country. Bootstrapped standard errors are shown in parentheses (250 resamples).

I Robustness Check Results

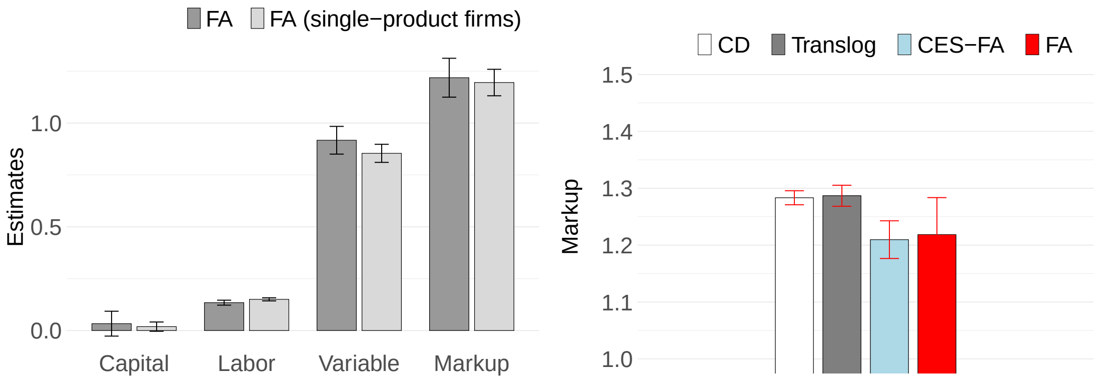
Figure OA-7: Elasticity Estimates and Markups from Robustness Exercises



(a) Quantity Production Function



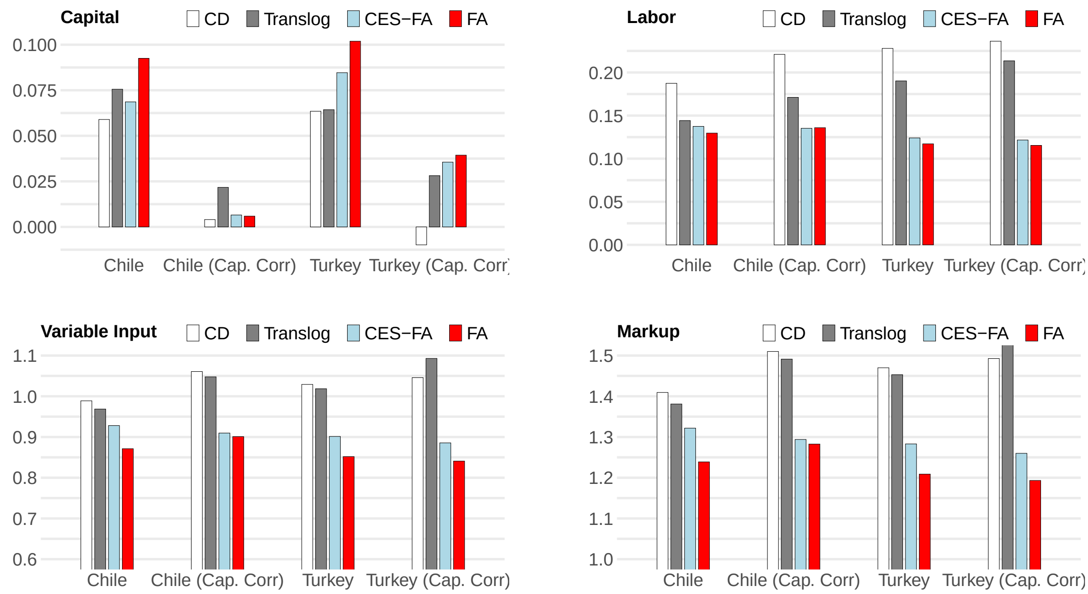
(b) Input Price Controls



(c) Single vs Multi Product Firm Controls

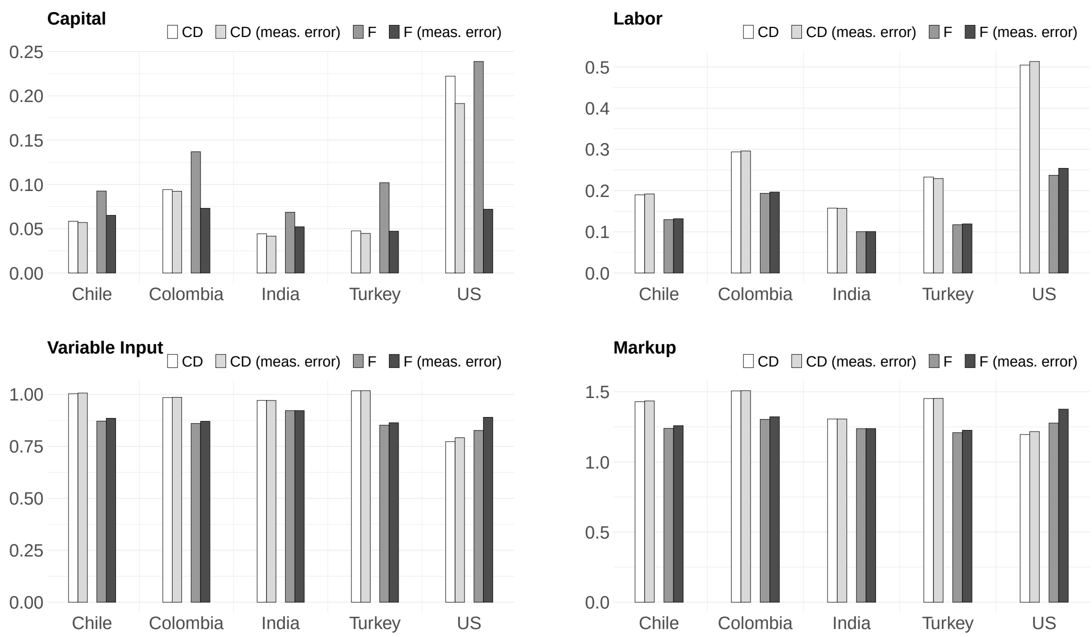
Notes: Each panel shows results from a different robustness check. The left panels compare estimates from the robustness-check specifications with the baseline empirical specification. The right panels show markup estimates from different production functions in each robustness check specification. Panel (a) uses production functions with output measured in physical quantities. Panel (b) controls for input prices in the production-function estimation. Panel (c) includes only single-product firms. See Online Appendix G for details of these robustness check specifications.

Figure OA-8: Comparison of Estimates with and without Capacity Utilization



Notes: This figure compares elasticity estimates from different production function specifications—CD, TR, CES-FA, and FA—for Chile and Turkey. It shows baseline estimates as well as estimates obtained after correcting capital stock for utilization. See Section Online Appendix G.3 for details on the estimation procedure.

Figure OA-9: Comparison of Estimates with and without Measurement Error



Notes: This figure presents results from the capital measurement error simulation exercise detailed in Online Appendix G.3. The white bars represent elasticity estimates obtained from the empirical procedure described in the paper. The grey bar represents the average of 250 elasticity estimates, each obtained by adding a random error to the capital stock.

References for Online Appendix

Acemoglu, Daron and Jensen, Martin Kaae. 2013. "Aggregate Comparative Statics". *Games and Economic Behavior* 81:27–49.

Ackerberg, Daniel A., Caves, Kevin, and Frazer, Garth. 2015. "Identification Properties of Recent Production Function Estimators". *Econometrica* 83 (6): 2411–2451.

Allcott, Hunt, Collard-Wexler, Allan, and O'Connell, Stephen D. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India". *American Economic Review* 106 (3): 587–624.

Anderson, Simon P., Erkal, Nisvan, and Piccinin, Daniel. 2020. "Aggregative Games and Oligopoly Theory: Short-Run and Long-Run Analysis". *The RAND Journal of Economics* 51 (2): 470–495.

Arkolakis, Costas and Morlacco, Monica. 2017. "Variable Demand Elasticity, Markups, and Pass-through". *Working Paper.*

Blundell, Richard and Bond, Stephen. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions". *Econometric Reviews* 19 (3): 321–340.

Imbens, Guido and Newey, Whitney. 2009. "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity". *Econometrica* 77 (5): 1481–1512.

Keller, Wolfgang and Yeaple, Stephen R. 2009. "Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States". *The Review of Economics and Statistics* 91 (4): 821–831.

Klenow, Peter J and Willis, Jonathan L. 2016. "Real Rigidities and Nominal Price Changes". *Economica* 83 (331): 443–472.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity". *Econometrica* 71 (6): 1695–1725.

Melitz, Marc J and Ottaviano, Gianmarco IP. 2008. "Market Size, Trade, and Productivity". *The Review of Economic Studies* 75 (1): 295–316.

Parenti, Mathieu, Ushchev, Philip, and Thisse, Jacques-François. 2017. "Toward a Theory of Monopolistic Competition". *Journal of Economic Theory* 167:86–115.

Raval, Devesh. 2023. "Testing the Production Approach to Markup Estimation". *Review of Economic Studies* 90 (5): 2592–2611.

Taymaz, Erol and Yilmaz, Kamil. 2015. "Estimating Plant-level Marginal Costs and Mark-ups in the Turkish Manufacturing Industry". *Middle East Technical University, ERC Working Paper.*

Production Function Estimation with Factor-Augmenting Technology: An Application to Markups

Mert Demirer

Supplementary Appendix - Not For Publication

A Descriptive Statistics of Production Datasets

Table SA-1: Descriptive Statistics - Chile

		Share (Sales)			Number of Plants		
ISIC	Industry	1979	1988	1996	1979	1988	1996
311	Leather Tanning and Finishing	0.17	0.19	0.20	1245	1092	983
381	Metal Products	0.04	0.04	0.04	383	301	353
321	Textiles	0.05	0.04	0.02	418	312	257
331	Repair Of Fabricated Metal Products	0.03	0.02	0.03	353	252	280
322	Apparel	0.02	0.02	0.01	356	263	216
	Other Industries	0.69	0.69	0.69	2399	1957	1873

Notes: Descriptive Statistics for Chile. Columns 3-5 show each industry's share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Columns 6-8 report the number of active plants in corresponding years. The last row provides information about industries not included in the sample.

Table SA-2: Descriptive Statistics - Colombia

		Share (Sales)			Number of Plants		
ISIC	Industry	1978	1985	1991	1978	1985	1991
311	Leather Tanning and Finishing	0.21	0.21	0.20	971	840	976
322	Apparel	0.03	0.03	0.03	666	862	842
381	Metal Products	0.04	0.04	0.03	593	478	534
321	Textiles	0.11	0.09	0.08	467	398	428
342	Cutlery, Hand Tools, and General Hardware	0.02	0.03	0.02	325	315	342
382	Laboratory Instruments	0.02	0.02	0.02	285	266	307
352	Farm and Garden Machinery and Equipment	0.06	0.07	0.09	287	257	305
369	Miscellaneous Electrical Machinery	0.03	0.04	0.03	299	257	267
356	General Industrial Machinery	0.02	0.03	0.04	197	252	341
	Other Industries	0.45	0.45	0.46	3893	3673	4001

Notes: Descriptive Statistics for Colombia. Columns 3-5 show each industry's share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Columns 6-8 report the number of active plants in corresponding years. The last row provides information about industries not included in the sample.

Table SA-3: Descriptive Statistics - India

		Share (Sales)			Number of Plants		
NIC	Industry	1998	2007	2014	1998	2007	2014
230	Other non-metallic mineral products	0.09	0.05	0.08	596	1056	1386
265	Measuring and testing, equipment	0.01	0.02	0.02	272	877	750
213	Pharmaceuticals, medicinal chemical	0.01	0.01	0.01	186	479	670
304	Military fighting vehicles	0.04	0.03	0.07	118	383	704
206	Sugar	0.06	0.04	0.04	271	363	431
	Other Industries	0.79	0.86	0.78	1172	2795	3510

Notes: Descriptive Statistics for India. Columns 3-5 show each industry's share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Columns 6-8 report the number of active plants in corresponding years. The last row provides information about industries not included in the sample.

Table SA-4: Descriptive Statistics - US

		Share (Sales)			Number of Firms		
NAICS	Industry	1961	1987	2014	1961	1987	2014
33	Manufacturing I	0.39	0.37	0.60	113	1092	752
32	Manufacturing II	0.51	0.53	0.25	84	392	222
31	Manufacturing III	0.10	0.10	0.15	36	138	104

Notes: Descriptive Statistics for the US. Columns 3-5 show each industry's share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Columns 6-8 report the number of active plants in corresponding years. The last row provides information about industries not included in the sample.

Table SA-5: Descriptive Statistics - Turkey

		Share (Sales)			Number of Plants		
ISIC	Industry	1983	1991	2000	1983	1991	2000
321	Textiles	0.16	0.13	0.16	1017	945	1803
311	Food	0.12	0.12	0.11	1261	1120	1061
322	Apparel	0.02	0.05	0.04	300	831	800
381	Metal Products	0.04	0.04	0.04	650	542	834
382	Machinery	0.05	0.06	0.04	532	482	683
383	Electrical-Electronic Machinery	0.04	0.03	0.04	413	523	639
356	Plastic Products	0.08	0.07	0.07	309	312	402
352	Pharmaceuticals	0.08	0.09	0.12	331	286	428
371	Motor Vehicles and Motor Vehicle Equipment	0.02	0.02	0.03	287	261	383
312	Beverage and Tobacco Product Manufacturing	0.05	0.06	0.07	263	218	250
	Other Industries	0.33	0.34	0.29	5100	5302	7033

Notes: Descriptive Statistics for Turkey. Columns 3-5 show each industry's share as a percentage of sales in the entire manufacturing industry for the first and last year, and at the midpoint of the sample. Columns 6-8 report the number of active plants in corresponding years. The last row provides information about the industries that are not included in the sample.

Table SA-6: List of Products for Quantity Production Function Estimation

Product Category	*Unit*	*Obs*	*Products Included (Code)* 2000-09 (ASICC)	2010-14 (NPCMS)
Brick Tiles	Tonnes	7500	29101 29102	3732001 3732007
Black Tea	Kilograms	6902	12211-5	239130-3 2391308
Rice, Parboiled Non-Basmati	Tonnes	6547	12311	2316107 2316202
Biri Cigarettes	Number of Cig.	5735	15323	2509001
Rice, Raw Non-Basmati	Tonnes	5057	12312	2316108 2316203
Shirts, Cotton	Number of Shirts	3515	63428 63428	2822203 2822299 2822406 2822408 2823499

Notes: This table presents the list of products that were used to estimate the quantity-based production function. The second column shows the unit of measurement, and the third column shows the number of firm-year observations for each product. The final two columns list the product codes that are included in each product category. The name of the products for each code can be found here.

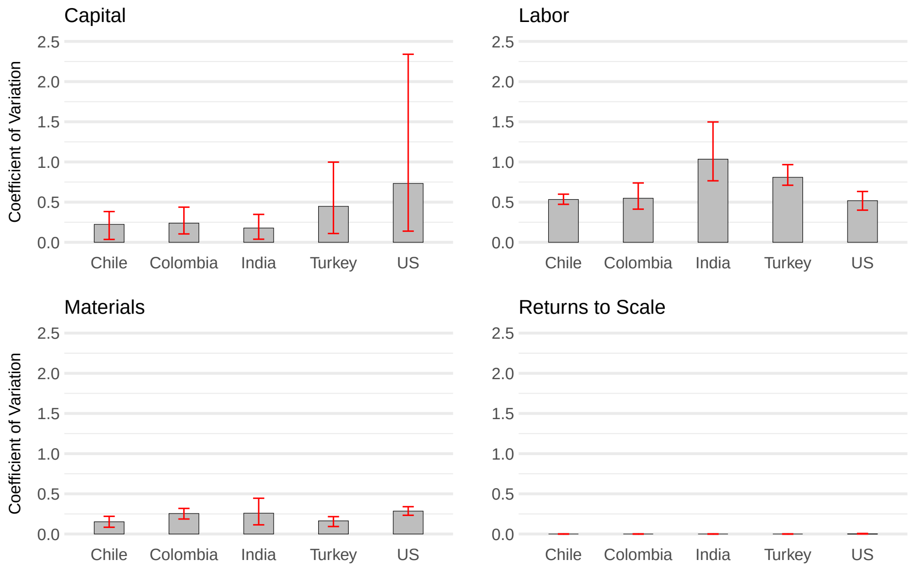
Table SA-7: Descriptive Statistics on Largest Input in Each Industry

Product	*Cost Share*	*CoV*	*Unit*	*Input Name*	*Code* 2001-09	*Code* 2010-14
Brick Tiles	0.62	1.99	Tonne	Clay, Common	21405	1540007
Black Tea	0.76	0.42	Kg	Tea (Green), Leaf	12202	162002
Rice, Parboiled Non-Basmati	0.88	0.32	Tonnes	Paddy (Excl. Seed)	12301	113200
Biri Cigarettes	0.81	1.05	Tonne	Kendu (Biri) Leaves	15318	2509002
Rice, Raw Non-Basmati	0.90	0.26	Tonnes	Paddy (Excl. Seed)	12301	113200
Shirts, Cotton	0.53	0.38	Metres	Fabrics, Cotton	63302	2669010

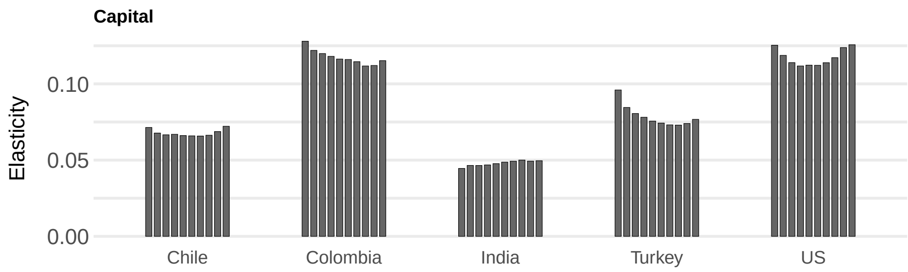
Notes: This table lists the largest inputs used in each product category. The second column shows the cost share of the input in intermediate input expenditures, and the third column presents the average CoV of each price in each year-industry. The final two columns provide the name of the input and the codes for two different time periods. The year 2000 is omitted because the largest input is "Other basic items (indigenous)" in almost all industries.

B Additional Figures and Tables

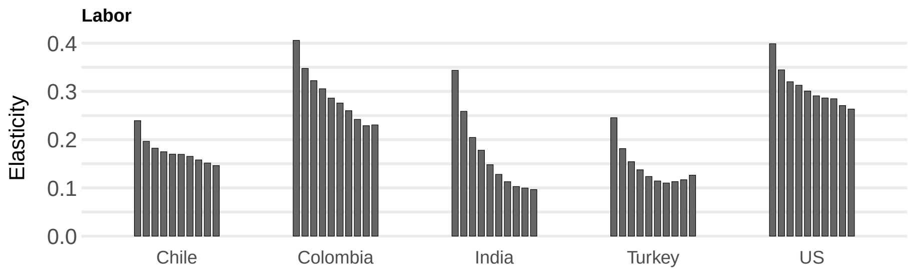
Figure SA-1: Heterogeneity in Output Elasticities (Nested CES)



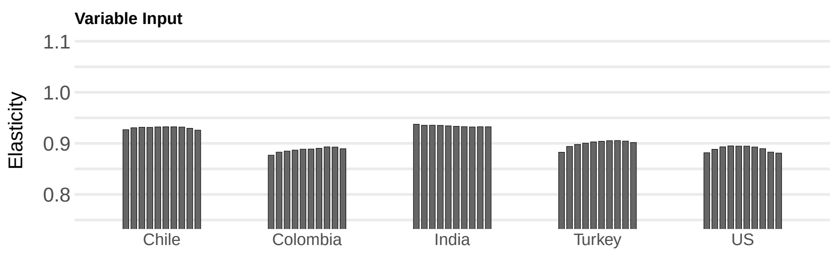
(a) Coefficient of Variation



(b) Capital Elasticity and Firm Size



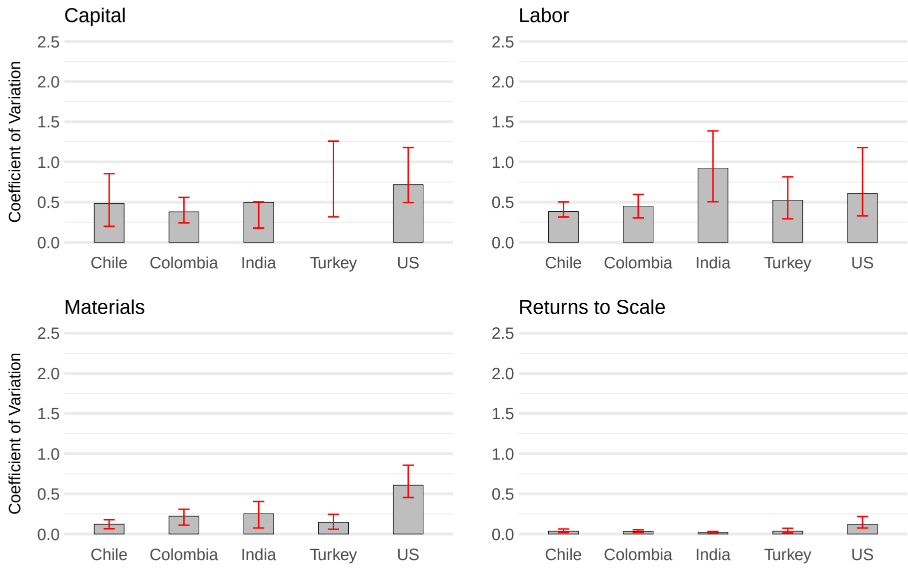
(c) Labor Elasticity and Firm Size



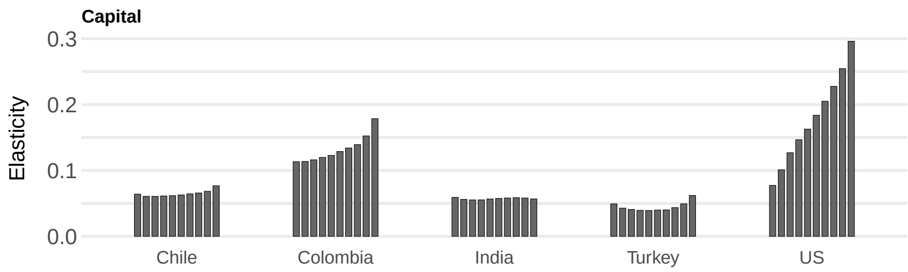
(d) Variable Input Elasticity and Firm Size

Notes: Panel (a) shows the average CoV of output elasticities estimated from the Nested CES specification. Panels (b-d) display average output elasticities by firm decile. For each country, the average elasticity for each decile within an industry-year is first estimated, and then these estimates are averaged across industry-year bins.

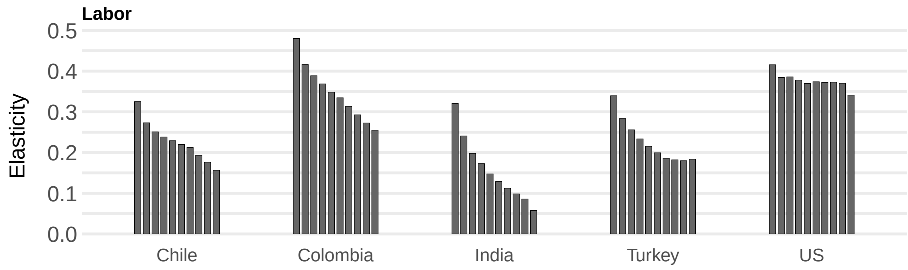
Figure SA-2: Heterogeneity in Output Elasticities (Translog)



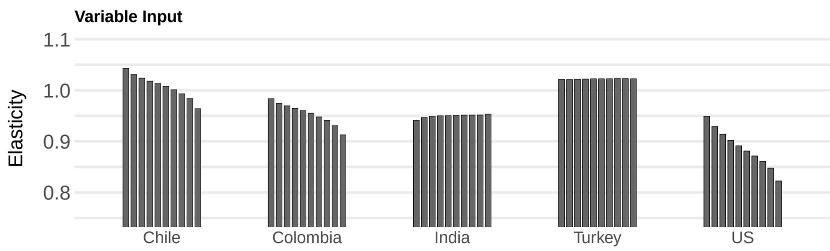
(a) Coefficient of Variation



(b) Capital Elasticity and Firm Size



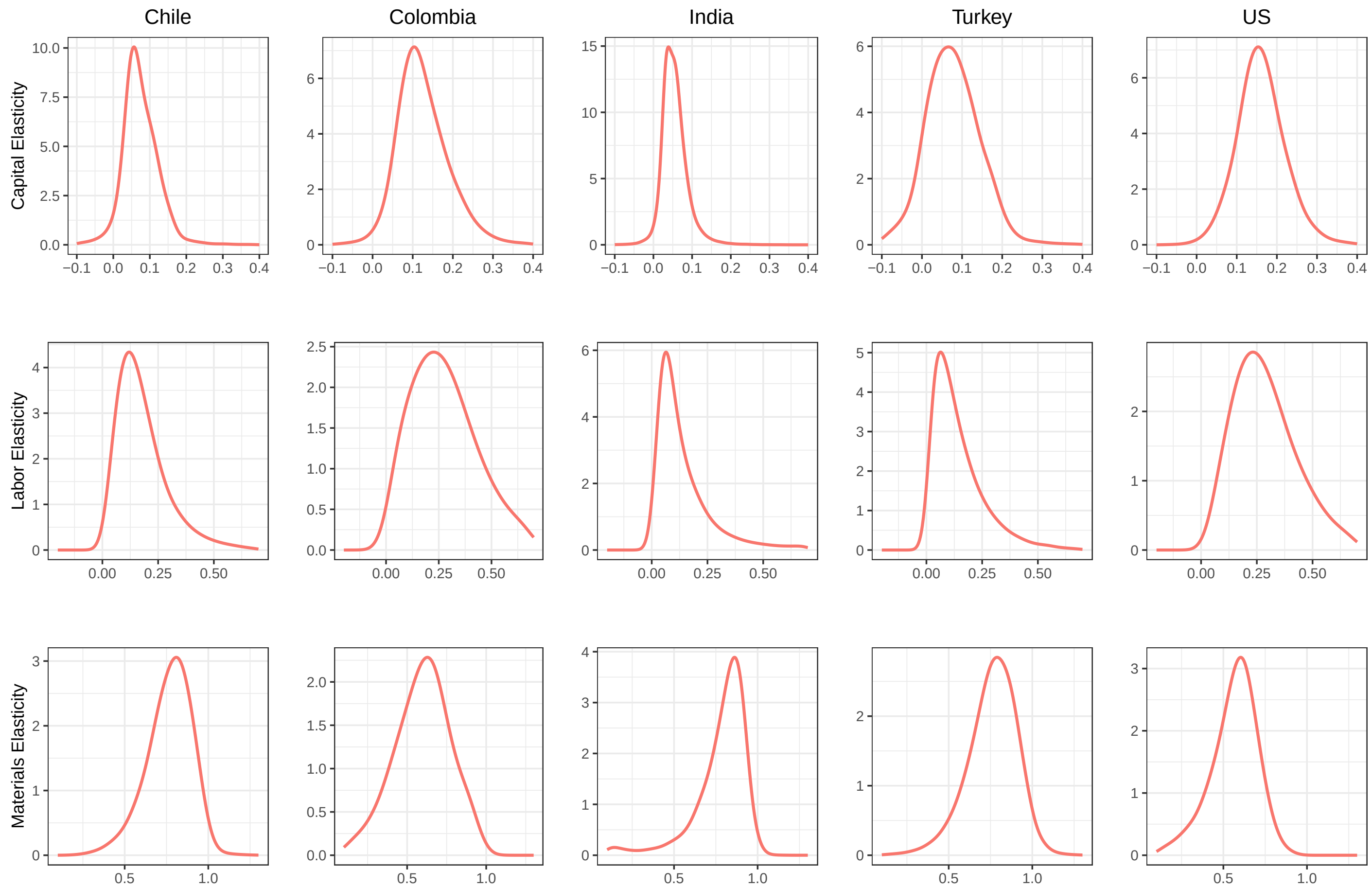
(c) Labor Elasticity and Firm Size



(d) Variable Input Elasticity and Firm Size

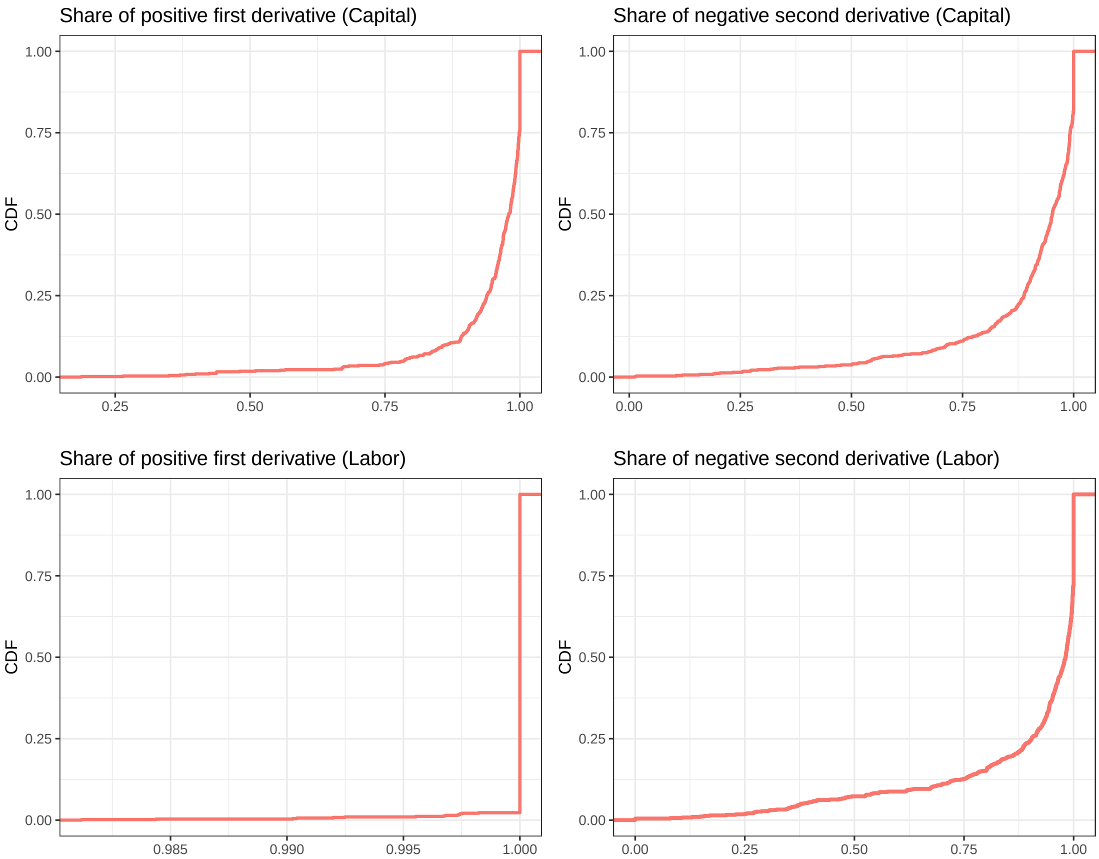
Notes: Panel (a) shows the average CoV of output elasticities estimated from the Translog specification. Panels (b-d) display average output elasticities by firm decile. For each country, the average elasticity for each decile within an industry-year is first estimated, and then these estimates are averaged across industry-year bins. In Panel (a), the average CV of capital elasticity in Turkey is implausibly high due to an outlier and therefore is out of the y-axis range.

Figure SA-3: Distribution of Elasticities Estimated using FA Method



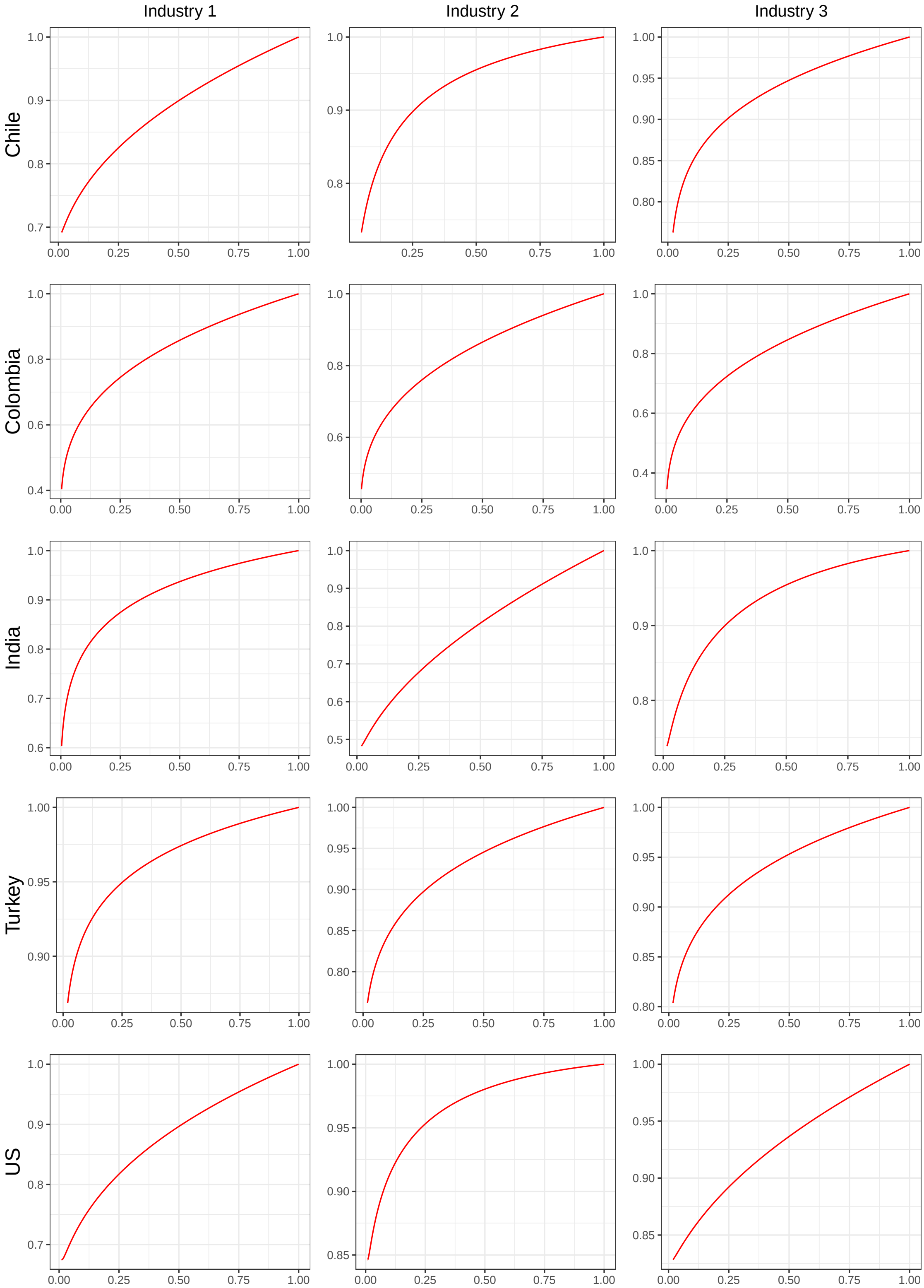
Notes: These figures display the distribution of capital, labor, and materials elasticity in each country estimated from the strongly homothetic production function with factor-augmenting productivity (FA). The unit of observation is a plant- or firm-year, depending on the country.

Figure SA-4: First and Second Derivatives of Production Functions



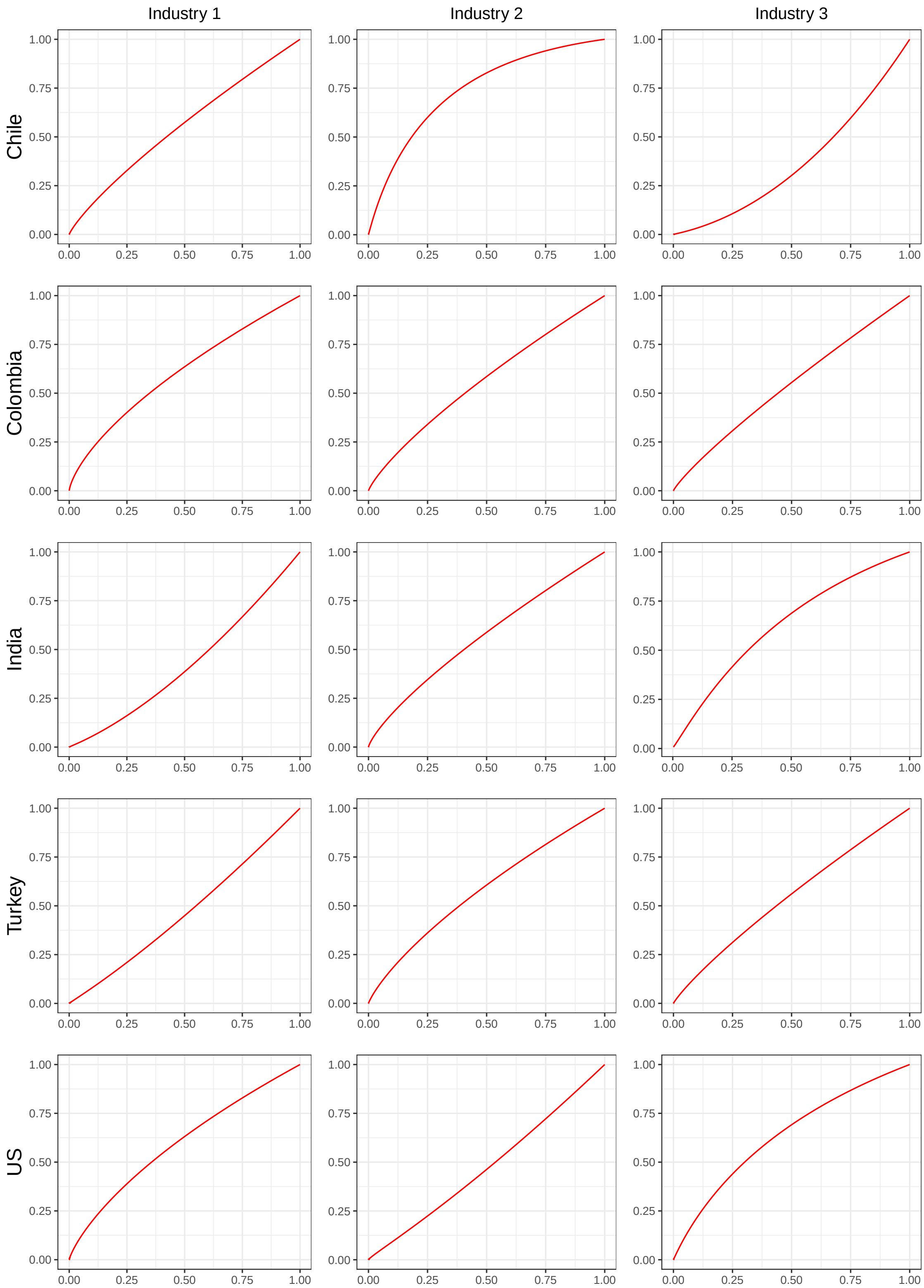
Notes: This figure displays the CDF of the proportion of firms with positive first derivatives and negative second derivatives with respect to capital and labor. In each CDF, the unit of observation is a country-year-industry, with a total of 601 observations. The first and second derivatives are evaluated at each firm's observed values of capital, labor, and materials. To give an example of how to interpret these plots, the top-right figure shows that in 90% of country-year-industry cases, less than 10% of the derivatives of the production function with respect to capital are negative.

Figure SA-5: Shape of the Production Function with Respect to Capital



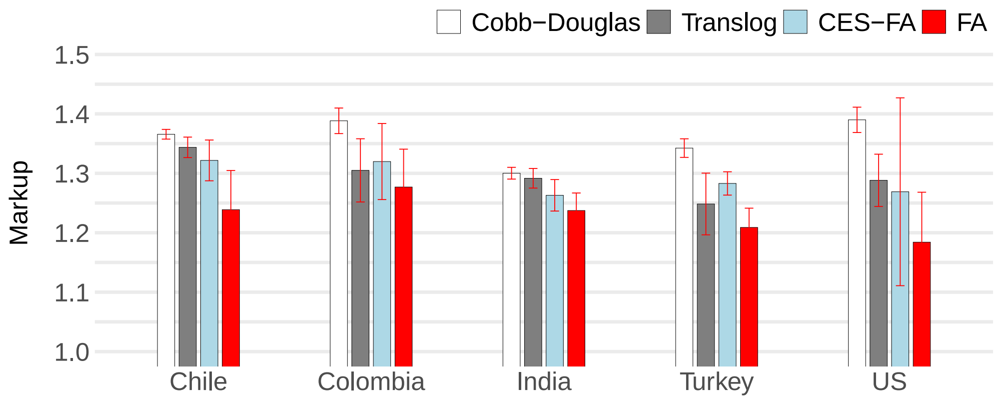
Notes: This figure displays the shape of the production function with respect to capital in the largest three industries of each country in the midpoint period of the sample. The values of other inputs are set to their median values observed in the sample. On the x-axis, the capital input moves from the smallest to the largest observed value in the sample, and the y-axis reports the corresponding change in output. Both axes are normalized to be between 0 and 1 to ease readability.

Figure SA-6: Shape of the Production Function with Respect to Composite Input



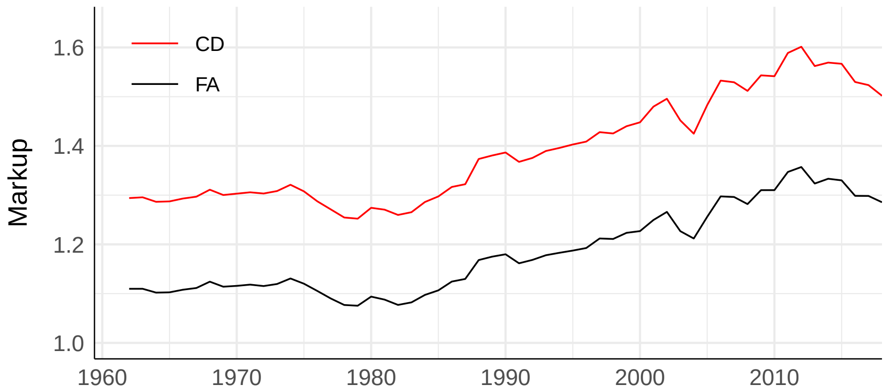
Notes: This figure displays the shape of the production function with respect to composite input h in the largest three industries of each country in the midpoint period of the sample. The values of other inputs are set to their median values observed in the sample. On the x-axis, the input moves from the smallest to the largest observed value in the sample, and the y-axis reports the corresponding change in output. Both axes are normalized to be between 0 and 1 to ease readability.

Figure SA-7: Average Markup Estimates Across Production Function Models (No Rolling Window)



Notes: Comparison of sales-weighted markups estimated using four production functions. Sales-weighted averages are calculated for each industry-year, then averaged across years. Error bars show 95% confidence intervals based on bootstrap. Due to the increase in the computational intensity, the standard errors are estimated using 50 resamples instead of 250 resamples as in the main text.

Figure SA-8: Sales-Weighted Markup Trend Over Time in US Manufacturing (No Rolling Window)



Notes: This figure compares sales-weighted markup trends in US manufacturing derived from a Cobb-Douglas (CD) production function versus a strong homothetic production function with Hicks-neutral and factor-augmenting productivity (FA).